Loschky, L. C., Hutson, J. P., Smith, M. E., Smith, T. J., & Magliano, J. P. (2018). Viewing Static Visual Narratives through the Lens of the Scene Perception and Event Comprehension Theory (SPECT). In A. Dunst, J. Laubrock, & J. Wildfeuer (Eds.), Empirical Comics Research: Digital, Multimodal, and Cognitive Methods (pp. 217-238). New York, NY: Routledge.

11 Viewing Static Visual Narratives through the Lens of the Scene Perception and Event Comprehension Theory (SPECT)

Lester C. Loschky, John P. Hutson, Maverick E. Smith, Tim J. Smith, and Joseph P. Magliano

1. Introduction

In everyday life, we are continuously presented with visual information, much of it in the form of stories or narratives. The visual narratives we consume come in many forms, from picture stories read by children to the comics read by adolescents and adults, to the advertisements, TV shows, and movies watched by people of all ages. While we have learned a great deal about reading and comprehending textual narratives over the last 100+ years of research (Rayner; McNamara and Magliano), we know far less about how we perceive and comprehend visual narratives (see Cohn). In this paper, we ask how moment-to-moment processing of visual narratives progresses, both in terms of perceptual and comprehension mechanisms. Importantly, how are these perceptual and comprehension mechanisms coordinated? We frame our discussion in terms of the Scene Perception and Event Comprehension Theory (SPECT) (Loschky et al. "Explaining"; "Scene Perception Applied"; "SPECT"), which we use as a lens to describe our research on visual narratives. The studies described used a particular set of visual narratives by Mercer Mayer, collectively known by many researchers as the "Boy, Dog, Frog" stories (Mayer, *Boy* and *One*) to test hypotheses generated by SPECT. A full account of the theory can be found elsewhere (Loschky et al., "SPECT"). As a small caveat for this volume on research on comics, we note that while the Boy, Dog, Frog (BDF) picture stories are visual narratives, and very similar to comics, they have a single picture per page rather than a multi-panel layout. Nevertheless, we feel that readers interested in comics perception and comprehension will have a similar interest in our studies. Likewise, our studies may also be relevant to readers interested in film, which shares similar demands on visual scene perception and sequential narrative comprehension, even though it differs radically from static visual narrative sequences in many critical ways (e.g., image motion, non-self-paced viewing, and audio).

SPECT is an integrative theoretical framework for understanding both visual narrative perception and comprehension (Loschky et al., "Explaining"; "Scene Perception Applied"; "SPECT"). It incorporates theories from the domains of scene perception (Henderson and Hollingworth, Wolfe et al., Oliva), event perception (Kurby and Zacks, "Segmentation"; Zacks et al., "Event"), and narrative comprehension (McNamara and Magliano; Gernsbacher, *Language*). SPECT has been developed to create bridges across these normally siloed research areas, as a framework for understanding visual narrative perception and comprehension from the onset of the first image in a visual narrative until recall of the entire narrative years later. To our knowledge, SPECT is the first theory to attempt such a grand synthesis across these disparate areas of research and theory. Although we only briefly outline SPECT here, its utility in developing and testing hypotheses within the BDF studies demonstrates the importance of developing a comprehensive theory of visual narrative perception and comprehension.

As shown in Figure 11.1, SPECT draws a key distinction between front-end and back-end mechanisms. Front-end mechanisms occur during single eye fixations, whereas back-end mechanisms occur across multiple fixations in working memory (WM) and long-term memory (LTM) over extended periods of time. Importantly, SPECT assumes that front-end and back-end mechanisms interact and influence each other bi-directionally. There are two primary front-end mechanisms that occur during each eye fixation: (1) information extraction from the image, and (2) attentional selection for where to send the eyes for the next fixation. These mechanisms are assumed to occur in parallel (Findlay and Walker; Laubrock et al.).

The front-end information extraction mechanisms include everything that happens *within a single eye fixation*, from the moment that light from the image first hits the retina until the recognition of a scene, person, object, or event, roughly 150–300 milliseconds (ms) later (Ramkumar et al.; Cichy et al.; Fei-Fei et al.), which enters into working memory (WM) in the back-end. Importantly, information extraction can either be *broad*, encompassing the entire scene and arriving at a holistic semantic representation of it, called *scene gist* (Oliva; Loschky and Larson), or *narrow*, including only information about a specific person, animal, object, or event (Cichy et al.; Glanemann). Broader (coarser) information is typically acquired more quickly than narrower (finer) information (Fei-Fei et al.; Loschky and Larson; Hegdé; Schyns and Oliva). For example, broader basic level scene category information (e.g., forest) can be extracted in less than a single eye fixation. Narrower basic level actions by a character (e.g., riding a turtle) require two fixations to extract (Larson). Thus, increasingly detailed information can be extracted on a fixation-by-fixation basis in the front-end, which is accumulated across multiple fixations in WM (Hollingworth and Henderson; Pertzov et al.) in the event model in the back end. The front-end attentional selection mechanisms occur

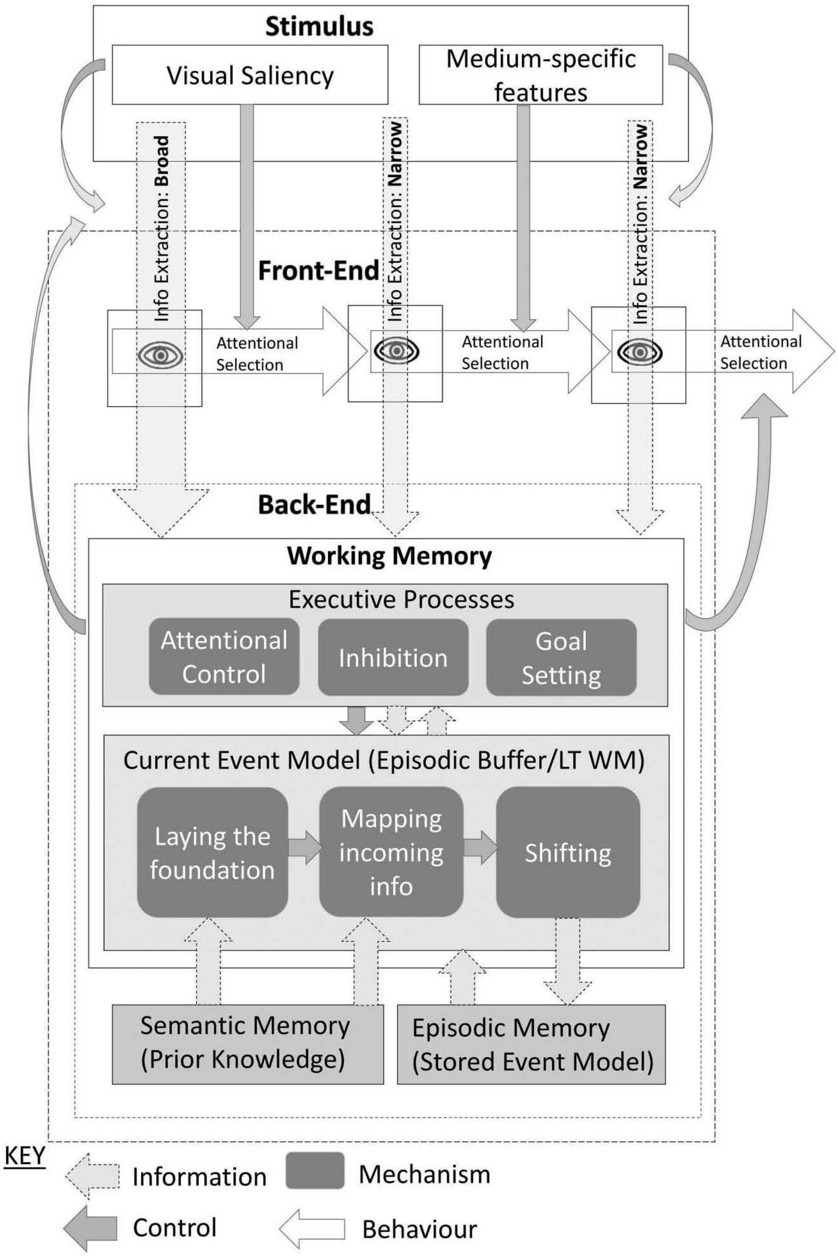

Figure 11.1 Box model of the scene perception and event comprehension theory (SPECT).

in parallel during each fixation, based both on the visual saliency of the stimulus features (Itti and Koch, Mital et al.), medium-specific features such as the organization of panels in a comic (Cohn), or the compositional features of edited shot sequences in movies (Smith, "Cinematic"), and back-end guidance (to be described in detail below).

There are several major classes of back-end mechanisms incorporated into our model: Executive mechanisms; mechanisms involved in creating and updating the current event model in WM; mechanisms in LTM, including previously stored event models in episodic LTM; and schemas, scripts, and other background knowledge accessed from semantic LTM. The *current event model* is the viewer's understanding of what is happening 'now' in the visual narrative, and is thus of particular interest in SPECT. Key information in the event model is in the form of event indices, including the event (i.e., meta-actions), entities (people, animals, objects), location, time, goals, and causal relationships (Zwaan and Radvansky; Magliano et al. "Aging"). There are three primary mechanisms involved: (1) laying the foundation for the current event model, (2) mapping new incoming information to it, and (3) shifting to create a new event-model (Gernsbacher, *Language*).

Laying the foundation is the process of creating a new event model from scratch. Take, for example, the top left image in Figure 11.2. Laying the

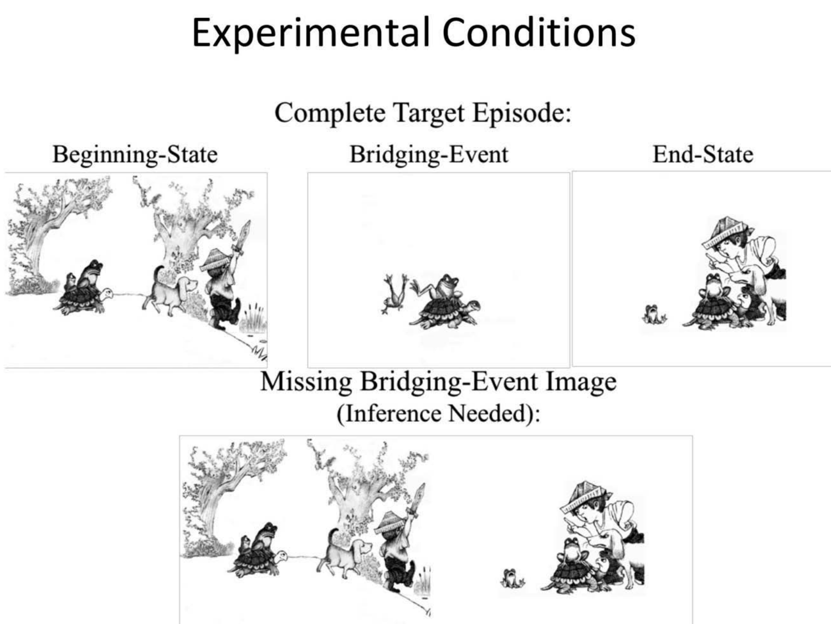

Figure 11.2 Experimental manipulation conditions for bridging inference. A complete 3-image target episode from Figure 11.2 is shown, including beginning state, bridging event, and end-state images. The missing bridging-event condition requires the viewer to infer the bridging event when they see the end-state image.

foundation begins with recognizing the gist of a scene (i.e., the spatial context, e.g., a forest), what sort of event is occurring (e.g., riding), and the main entities involved, including who is the agent and the patient (e.g., two frogs on a turtle). These event indices (space, action, agent(s), patient) can generally be extracted within one to two fixations (Larson; Larson and Loschky; Glanemann). This information activates associated knowledge from semantic long-term memory, which informs the new event model, producing expectations for upcoming information (Eckstein et al.; Torralba et al.).

Mapping further information to the current event model occurs as long as it is coherent. This involves adding information about other event indices, such as other entities (people, animals, objects), their goals (e.g., the big frog wants to dispose of the little frog), and causal relationships (e.g., the little frog is crying because it was kicked off of the turtle). Shifting to create a new event model occurs either when predictions of the current model fail (Zacks et al., "Event"), when incoming information is incoherent with the current model (Zwaan et al.), or when one or more important event indices change (Magliano et al., "Aging"; Huff et al). When this happens, the current event model is stored in episodic LTM, becoming a *stored event model.* Stored event models in episodic LTM can then be used to inform the new event model.

Finally, executive processes in WM involve conscious control of processing through goal setting (e.g., being able to summarize the visual narrative), attentional control (e.g., consciously searching for specific information in the narrative), and inhibition (e.g., consciously ignoring information deemed irrelevant). Some of these executive processes can be effortful (e.g., goal setting and execution of strategies to achieve those goals), and may not be very involved in normal, seemingly effortless comprehension of visual narratives. However, they can exert a strong influence on comprehension to the degree that viewers think of using them, and have the executive resources available to do so (Moss et al.).

An important contribution of SPECT is that it enables a principled investigation of the interactions between the front-end and back-end mechanisms that have not been studied in scene perception and comprehension before due to artificial boundaries between the research areas of scene perception, event perception, and narrative comprehension. For example, SPECT provides important details to help understand how specific front-end mechanisms, such as gist recognition as a type of broad information extraction, influence specific back-end mechanisms, such as laying the foundation of the event model. SPECT also proposes how specific back-end mechanisms, for instance, those involved in mapping incoming information to the current event model in WM, can affect specific front-end mechanisms, such as attentional selection. While these sorts of questions have been investigated in the area of reading, they have rarely been investigated in the areas required by visual narrative comprehension (but see Foulsham et al.).

2. What Is the Nature of Back-End Mechanisms in 'Reading' Picture Stories?

SPECT assumes that the back-end processes involved in the construction of the event model are general cognitive mechanisms, which extend beyond the language-based modality to include visual narratives. This assumption has been supported by the data from our studies, which show evidence consistent with three stages of building the current event model: laying the foundation, mapping, and shifting (Hutson et al; Magliano et al., "Filling," "Aging," "Relative"; Smith et al., "Bridging" and "Laying"). There is a growing body of research that has been conducted on each of these back-end mechanisms in the context of visual narratives.

Two distinct processes that support the three phases of building the current event model are event segmentation and bridging inference generation (Magliano et al., "Reading"). Event segmentation is the process of detecting boundaries between events, or between sub-events that comprise larger episodes. Segmentation is either incremental or global (Kurby and Zacks, "Starting"). When a new event begins, such as when the behavior of characters suggests that a prior goal has been completed, the old event representation is abandoned and a new one is created (i.e., a global shift) (Magliano et al. "Aging"). Changes in the behavior of characters engaged in a goal plan that move the agent toward goal completion indicate incremental boundaries. Figure 11.3 shows eight pictures from the beginning of a story, three of which have dashed lines around them, indicating they

Figure 11.3 Illustration of event segmentation task in a BDF story (adapted from Mayer's "One Frog Too Many"). After participants read the entire story, they saw thumbnails of all the images in sequential order. Their task was to click on each picture when they thought a change in the story situation occurred. Normative event segmentation for this story is indicated in this illustration by dashed lines around the most frequently segmented images.

were normatively segmented. The three segmented pictures appear to be segmented at different levels of granularity. The third segmented picture appears to be the conclusion of a coarse, global event and the initiation of a new goal (e.g., the boy, dog, frogs, and turtle go for a walk in the woods). The first two segmented pictures appear to be sub-events within a larger episode (of receiving a present and ensuing interaction among the characters), with the sub-events being the opening of the present and discovery of a new frog. Segmenting events in this way enables the reader to treat each sub-event or larger event as a single chunk of information, thus saving processing resources and improving comprehension when done well (e.g., Kurby and Zacks, "Segmentation"). Incremental updating may signal the initiation of mapping processes because the boundaries reflect events that are part of the coarser event. Global boundaries should initiate shifting to a new event model. Inferences support mapping (Gernsbacher, *Language*), and therefore incremental boundaries may be a signal that inferential processes that support mapping should be engaged. However, to our knowledge, this claim has yet to be directly tested.

3. The Mechanism of Laying the Foundation: The Role of Front-End Processes

When someone sees the first image in a visual narrative, according to SPECT, they must lay the foundation for a new event model. Research has found that when reading a textual narrative, reading times for the first sentence are longer than for subsequent sentences (Haberlandt and Graesser). The same is true when viewing short comic strips (Foulsham et al.). In studies using the BDF stories, results have shown that the longest viewing times of any image are on the very first image in a story (Gernsbacher, *Memory*; Smith et al., "Laying"), which is similar to what has been found for first sentences of a textual story episode (Haberlandt et al.). According to SPECT, this is due to the process of laying the foundation for the new event model. Importantly, SPECT goes into some detail about how front-end processes lay the foundation. As noted earlier, SPECT argues for two chief front-end mechanisms that occur in parallel: *information extraction* and *attentional selection*. In laying the foundation for a new event model, on the very first fixation the information extraction process plays a critically important role. On that first fixation, the viewer perceives holistic semantic information about the scene in a coarse-to-fine manner: going from its subordinate level scene category (e.g., an outdoor scene), to its basic level category (e.g., a forest). This information activates expectations for where important information is in a scene, which affects attentional selection on that first fixation (Eckstein et al.). Recognizing the spatial context of the event (e.g., a forest) is a foundational event index for the new event model. Following that first fixation, the first eye movement in an image is usually made to something highly informative, particularly a person (if there is one), and by the end of the second fixation, the viewer will have recognized what that person is doing (i.e., the basic level action category, e.g., riding; Larson; Larson and Loschky). Having recognized the event indices of the spatial context, an entity (e.g., a potential protagonist), and the entity's action (the core event index), the viewer is well on their way to having laid the foundation for the new event model, and all within the time span of two eye fixations. Importantly, the three back-end event model mechanisms of laying the foundation, mapping, and shifting occur cyclically. Thus, after shifting to create a new event model, the viewer begins the cycle again by laying the foundation for the next event model.

When viewers look at a new image, their eye movements progress through two processing stages: (1) the *ambient mode*, characterized by longer saccade lengths and shorter fixation durations as viewers begin by exploring the image to figure out where the major scene constituents are (Pannasch et al.; Smith and Mital); and (2) the *focal mode*, characterized by shorter saccade lengths and longer fixation durations, as viewers extract detailed information from specific scene constituents that are of interest to them. A prediction of SPECT is that viewers will go through the ambient-to-focal stages of processing at event boundaries, with the ambient mode used in laying the foundation of the new event model. In our own studies with the BDF stories, we have found that saccade lengths were longer and fixation durations were shorter on images identified as event boundaries, consistent with a switch to the ambient mode at boundaries (Smith et al., "Laying"). Specifically, both saccade lengths and fixation durations showed significant interactions between boundary/non-boundary and viewing time on an image. For boundary images, mean saccade lengths remained at a constant 5 degrees of visual angle over the course of viewing, but for non-boundary images they showed roughly a 20% decrease (from 5 to 4 degrees). For boundary images, mean fixation durations remained shorter, increasing 22% more slowly over the first twenty seconds of viewing (going from 200 to 350 ms) than for the non-boundary images (going from 200 to 450 ms). These differences suggest that viewers remained in the ambient mode for longer while viewing an image identified as an event boundary than when they viewed non-boundary images. The latter results are consistent with the SPECT-generated hypothesis that at the beginning of a new event, viewers explore an image in the ambient mode to lay the foundation for the new event model. In terms of SPECT's two primary front-end mechanisms, the longer saccades would be evidence of a change in attentional selection and the shorter fixations would be attributed to a change in information extraction, both of which would relate to changes occurring in the current event mode (namely, shifting and laying the foundation). Nevertheless, saccade lengths and fixation durations, while generally independent, can show interdependencies (Findlay and Walker), and so further studies are needed to test these speculations.

4. The Mechanisms of Mapping and Shifting in the Context of Picture Stories

After having laid the foundation, according to SPECT, the viewer continues to map new incoming information to the current event model. New information is incorporated into the current event as long as the new information is coherent, does not change radically, and fits with expectations. However, if there is a lack of coherence, such as when there are changes to the various event indices of the current event model (new people, places, or goals), mapping stops and a shift occurs. This shift marks a global boundary at the end of the current event and the start of the next one.

Based on prior research showing that viewers perceive event boundaries when there are situational shifts in space, time, characters, goals, and causality, Magliano et al. ("Aging") conducted a study exploring the similarity in event segmentation across visual and text-based narratives using the BDF stories. Participants either viewed the BDF picture stories or read text-based versions of those stories written by the experimenters, such that they conveyed content as close as possible to that in the pictures. They also performed a content analysis of the original illustrated stories to determine when the pictures reflected changes in the event indices of time, space, characters, or goals. While viewing the picture stories or reading the texts, participants also carried out an event segmentation task in which they decided whether a story unit (picture or sentence) was the start of a new event (e.g., Zacks et al., "Segmentation"; see Figure 11.3). Magliano et al. ("Aging") used the situational change content analysis to determine if changes in the event indices were correlated with segmentation judgments, and to determine if there were differences in these relationships as a function of modality (i.e., picture vs. text).

Magliano et al. ("Aging") found that judgments of event boundaries were significantly correlated with all changes in event indices (i.e., shifts in situational continuity), which is consistent with SPECT in terms of the relationship between the mapping and shifting mechanisms. If there is a large enough change in situational continuity (i.e., changes in event indices), it signals a lack of coherence, which leads to shifting, marking an event boundary. They also found that the magnitude of the correlations was similar across the visual and text-based versions, which indicates that segmentation is similar across visual and text-based narratives. However, there was one interesting difference between the modalities: Viewers' event segmentation decisions were less influenced by changes in the time and space event indices in the picture stories than in the text versions. While this particular result is in need of replication, one possible explanation for it is that viewers of visual narratives give greater weight to changes in goals than to changes in time and space when judging event boundaries, which has also been found for viewers of a film narrative ("The Red Balloon"; Magliano and Zacks).

Why do readers of text weight spatio-temporal changes more heavily than readers of visual narratives? Magliano et al. ("Aging") speculated that this may have been an artifact of the materials and their translations. Many of the narrative events take place in similar settings (e.g., different parts of a forest, different parts of a park). Therefore, the products of the front-end mechanism of broad information extraction (i.e., scene gist processing) could have been very similar across pictures that depicted characters changing locations in these settings, thus attenuating the perceived shifts in space. However, these changes were explicitly marked with temporal adverbs (e.g., a little while later, the next day) and spatial prepositions (e.g., into the woods, into the open field, on the log) in the text-based adaptation. In the context of SPECT, the results of Magliano et al. ("Aging") confirm the importance of segmentation to both the mapping and shifting mechanisms, and that changes in situational continuity signal changes in the current event model (see also Kurby and Zacks, "Starting").

There are various processing costs that accompany shifting to create a new event model. In reading, when incoming information is coherent with the current model, it is readily mapped onto the event model (Zwaan and Radvansky), requiring minimal processing resources. However, when there is a change on one or more event indices, or an event boundary is passed, processing becomes much more intensive, producing slower reading times (Zacks et al., "Segmentation") and longer fixation durations (Swets and Kurby). As part of this process, the current event model is stored in episodic LTM when a shift occurs, which is analogous to processes involved in *sentence wrap-up*. The latter have also been found to produce longer readings times at the end of sentences and clauses (Just and Carpenter). In our studies using the BDF stories, visual narrative 'readers' viewed pictures identified as boundaries for longer than non-boundary pictures (Smith et al., "Laying"). Given that SPECT distinguishes front-end mechanisms of information extraction and attentional selection, an interesting question is whether these longer viewing times are due to viewers' longer fixation durations (more intensive information extraction), more fixations (more rapid shifts of attention), or both. As described above, in an eye movement study we found that fixation durations on images identified as boundaries were shorter, not longer, as compared to non-boundaries (Smith et al., "Laying"). In contrast, more fixations were made on those images. The higher number of fixations suggests that the back-end processes of shifting and laying the foundation are associated with greater exploration of boundary images as reflected by changes in the front-end processes of information extraction and attentional selection. An important challenge for further research will be to firmly distinguish effects due to event wrap-up processes versus those that are due to laying the foundation for the next event.

5. The Mechanisms of Mapping and Shifting: Mapping and Bridging Inference Generation in Picture Stories

As noted earlier, whether or not new incoming information from the front-end is mapped onto the current event model depends on how well that new information coheres. However, visual narratives, like textual narratives, cannot show everything in their narrative worlds. Instead, only information that is crucial to the narrative is shown, which creates gaps in the narrative world that the visual storyteller assumes will not create comprehension problems for the viewer. Thus, when the viewer of a visual narrative is faced with such a narrative gap, they must decide whether it creates a coherence gap sufficient to warrant shifting to create a new event model, or if a reasonable inference can bridge the gap and maintain the coherence of the current event model (Graesser et al.). Thus, the mapping mechanism in SPECT crucially involves generating bridging inferences to fill the gaps between explicitly shown events, as has been consistently demonstrated with text materials (Graesser et al.).

In the context of this chapter, bridging inferences are important for making connections between panels and pictures in visual narratives. Consider Figure 11.2, which depicts a three-picture sequence from one of the BDF stories. The first panel shows the boy with his pets, including a big frog and a little frog, riding on the back of a turtle. As shown in Figure 11.3, we learn that the boy got a new little pet frog, of whom his older and bigger frog is jealous. Thus, in the second picture, the big frog kicks the little frog off of the turtle's back, and in the third picture the boy scolds the big frog. However, what would happen to one's comprehension of this episode if the second (middle) picture were missing? The viewer would have to infer why the little frog was on the ground crying while the boy was scolding the big frog in order to understand how that end-state picture is related to the current event model. Readers routinely make these bridging inferences to support constructing a coherent event model (an event model in which story constituents are semantically connected).

Magliano and colleagues (Magliano et al., "Relative," "Filling") showed that, indeed, viewers generate bridging inferences to connect pictures when processing visual narratives. Participants viewed the six BDF stories, and picture-viewing times were collected. An assumption was that viewing times for picture sequences where bridging inferences were needed would be longer than where they were not. To create this situation, we identified 24 three-picture sequences like the one shown in Figure 11.2, which consisted of a beginning state, a bridging event, and an end state. We then manipulated whether or not the bridging-event picture was present. In a pilot study, we had participants think aloud following the end-state pictures. We found that, consistent with a counterintuitive prediction, viewers were more likely to mention the bridging-event actions when the bridging-event pictures were *missing* than when they were present (because the bridging event would be more active in their working memory if they had needed to infer it). This suggests that when viewers read visual narratives silently, they infer bridging events when they are missing (Graesser et al.). Also consistent with our predictions, we found that viewing times for the end-state pictures were longer when the bridging-event pictures were absent (M = ~2,800 ms) than when they were present (M = ~2,450 ms). This prediction was based on research showing that sentence-reading times are longer when viewers need to generate bridging inferences (e.g., Clark). Together with the think-aloud data, our results provided converging evidence that viewers do indeed regularly generate bridging inferences online to map across gaps between pictures in visual narratives (c.f., Cohn and Wittenberg). This is also further evidence in favor of the claim that back-end mechanisms operate similarly in visual and textual narratives.

An interesting question is whether and how bridging inference generation processes relate to event segmentation. According to SPECT, this issue is conceptualized in terms of the relationship between the mapping and shifting mechanisms. If a gap in narrative coherence occurs and the coherence can be maintained by generating a bridging inference, then there is no need to shift to create a new event model. Thus, viewers would not be expected to give a segmentation response. Conversely, if the gap cannot be bridged by an inference or if the gap is judged to be too large, then there will be a shift to create a new event model, and the viewer will make a segmentation response.

We have recently investigated this question in an event segmentation study with the BDF stories (Smith et al., "Bridging"). Using the same self-paced viewing time paradigm as Magliano et al. ("Relative"), we manipulated the presence/absence of the bridging event images. After viewing each of the stories, participants identified images in the story that they felt signified a change in the story's situation (i.e., an event boundary) (Figure 11.3). Interestingly, participants were more likely to identify the end-state image as a boundary when the bridging event was absent than when it was present. This result is consistent with the idea that viewers were reacting to a perceived lack of coherence. However, we already presented evidence consistent with the idea that viewers generated bridging inferences when pictures were removed (e.g., Magliano et al., "Relative"), which allowed them to maintain coherence. As such, they would not be expected to make a segmentation response. The result showing increased event segmentation when missing the bridging action presents a theoretical puzzle. One possibility is to assume that our results are due to a mutually exclusive mixture of different participants making one or the other response (inference or a shift), but not both. Alternatively, character goals are still maintained when the bridging event is missing in this example. Therefore, these judgments may instead reflect incremental boundaries. As mentioned earlier, incremental boundaries may be a signal that the coherence break could be resolved via mapping processes. These are two testable hypotheses.

6. How Back-End Mechanisms Affect the Front-End Mechanism of Attentional Selection in Picture Stories

A unique contribution of SPECT is that it allows us to test novel hypotheses about the interactions between front-end and back-end mechanisms, which are assumed to be bidirectional (Figure 11.1). Thus, it is assumed that the front-end mechanisms of information extraction and attentional selection influence the back-end mechanisms involved in creating the current event model. That assumption forms part of most theories of narrative comprehension without comment. The finding presented above (that viewers remain in the ambient mode of processing longer while viewing images identified as boundaries) supports this assumption (Smith et al., "Laying"). On the other side, SPECT assumes that back-end mechanisms involved in creating the current event model influence moment-to-moment processes involved in information extraction and attentional selection in the front-end during eye fixations. These predictions are very novel within the areas of scene perception, event perception, and the comprehension of visual narratives and comics (see also Foulsham et al.).

Hutson, Magliano, and Loschky tested the assumption that front-end information extraction and attentional selection processes are sensitive to the back-end mapping process of inference generation. As presented above, Magliano et al. ("Relative," "Filling") showed that generating a bridging inference during picture-story viewing increased end-state image-viewing times. Using the same manipulation of bridging event presence or absence to induce inference generation, Hutson et al. measured viewers' eye movements to test the influence of the back-end on the front-end. Two possibilities could account for the increased viewing time during inference generation (Magliano et al., "Relative," "Filling"). Increased viewing times can be accounted for by eye movements in terms of increased fixation durations and/or more fixations. The computational load hypothesis proposed that viewers under the greater computational load of generating the inference would produce longer fixation durations (Just and Carpenter). The competing visual search hypothesis proposed viewers would search the scene for information relevant to making a bridging inference, thereby producing more fixations.

Surprisingly, there was essentially no difference in fixation durations between the bridging event-present and -absent conditions. There was, however, a significant 22% increase in the number of fixations produced, with viewers on bridging event-absent trials making two additional fixations on the end-state image compared to bridging event-present trials (approximately eleven vs. nine fixations, respectively). These results support the visual search hypothesis. This raises follow-up questions regarding what viewers fixated during the additional fixations. The inferential-informativeness hypothesis proposed that viewers needing to generate a bridging inference to maintain coherence between pictures would preferentially fixate regions that were more informative for drawing the inference.

To measure the inferential-informativeness of scene regions, we ran a rating study with new participants to identify inference relevant regions. These new participants were fully informed about the bridging event-present/event-absent manipulation in the previous eye-tracking experiment, and were asked to click the areas of the end-state scene that they thought would be relevant for inference if a participant had not seen the bridging event image. We then used these click locations to quantify the inferential-informativeness of image regions and found that for bridging event-absent trials, viewers were more likely to look at inference-relevant regions than in the bridging event-present trials. Thus, when participants needed to draw an inference, they used additional eye movements to pick up information relevant to generating that inference. For SPECT, this result shows that the back-end mapping process of bridging inference generation has an impact on front-end attentional selection.

Importantly, support for the visual search hypothesis shows a potential departure of comprehension in scenes from text, which supports the importance of SPECT. When there is a break in coherence, the pictures available allow for a visual search. In text, regressive eye movements could be considered analogous to visual search of a scene. However, in text, when readers need to generate a bridging inference, they typically don't use regressive eye movements to search for the relevant information for drawing the inferences. Rather, they are more likely to rely simply on activating knowledge in LTM (Singer and Halldorson), producing longer gaze durations on the target items (Myers et al.; O'Brien et al.). This asymmetry in the likelihood of using different processes to generate an inference between textual and visual narratives is likely due to the influence of medium-specific features of the stimulus on attentional selection (see Figure 11.1, top right). For example, the essential spatiotemporal linearity of attentional selection across multiple fixations while reading text (i.e., left to right in most languages) stands in contrast to the much less constrained spatiotemporal dynamics of attention in an image, even if that image is embedded in a sequential visual narrative.

7. Characteristics of the Stored Event Model in Long-Term Memory in Picture Stories

According to SPECT, when a viewer shifts to create a new event model, they first store the current event model in LTM. This instance then becomes part of a linked set of stored event models that represent the entire visual narrative, which can be recalled later, for example when retelling the story to a friend. SPECT assumes that the set of stored event models for a given narrative are structured around the goal episodes of characters for both textual and visual narratives (Baggett). We have investigated this issue in terms of understanding the implications of generating inferences on LTM for explicit versus inferred content. It is well documented that memory for narrative texts becomes distorted over time, such that memory for explicit content becomes weaker and memory for inferred content becomes stronger (e.g., Schmalhofer and Glavanov). Magliano and colleagues ("Filling") show that bridging inferences that connect pictures in a visual narrative distort LTM relatively quickly.

Magliano et al. ("Filling") had participants view the BDF stories and manipulated whether the beginning-state, bridging-event, and end-state pictures of the target episodes of Magliano et al. ("Relative") where shown. We measured participants' picture-viewing times during viewing, and after participants had viewed all six stories, gave them a picture recognition memory task. In this study, we measured viewing times for the pictures following beginning-state, bridging-event, and end-state pictures. Like Magliano et al. ("Relative"), we assumed that finding longer viewing times on those pictures where the preceding picture was missing indicated that the missing event was inferred. Viewing times were longer when the beginning- and bridging-state pictures were missing, but this was not the case for end-state pictures, suggesting that viewers were less likely to infer end-states. Importantly, performance on the recognition memory task showed that participants were also more likely to falsely remember having seen missing beginning-state and bridging-event pictures than end-state pictures (Magliano et al. "Filling"). These results suggest that generating bridging inferences distorts memory for the content of visual narratives. In terms of SPECT, this shows the workings of processes within the stored event models in LTM.

8. Conclusion

We all experience visual narratives in many different forms, including, but not limited to, comics, picture stories, and movies. While the current paper has focused on studies of picture stories as one form of visual sequential narrative, SPECT is equally relevant to other visual narrative formats such as comics and film. There are important differences between each of these formats, which are incorporated into SPECT in terms of medium-specific features (Figure 11.1: top right). These include conventions for reading the multi-panel page layout of comics (Cohn) and the perception of continuity as affected by a knowledge of editing conventions in film (Smith, *Editing*). Both of these cases would be predicted to not only influence the front-end process of attentional selection, but may also influence the back-end process of mapping incoming information onto the current event model. An important direction for further research is to test such predictions of SPECT in a broader effort to elucidate how these differences in visual narrative format between picture stories, comics, and film influence the front-end and back-end processes proposed by SPECT.

The theory and practice of creating visual narratives has far outpaced their empirical study. SPECT introduces a novel theoretical framework for those interested in the growing field of empirical research in visual narratives (see also Cohn for an alternative, but not mutually exclusive framework for comics). SPECT bridges well-developed theories on scene perception, event cognition, and narrative comprehension, which have thus far been compartmentalized. By providing an integrative framework encompassing these three areas, SPECT offers testable predictions about the bidirectional relationship between the front-end mechanisms of information extraction and attentional selection, and the back-end mechanisms in WM and LTM involved in constructing the current and stored event models.

One novel aspect of SPECT, which makes it a useful theoretical framework for those interested in visual narratives, is that it relies on well-established comprehension mechanisms identified in theories of reading comprehension, and generates predictions for how they operate in visual narrative comprehension. Importantly, while the comprehension processes themselves are based on general cognitive mechanisms, they may require and rely on different front-end mechanisms given the unique characteristics of the visual scene stimuli. As such, SPECT allows for tests of classic comprehension mechanisms (e.g., laying the foundation, shifting, and mapping). At the same time, SPECT asks how well-established perceptual and attentional mechanisms that have been identified in theories of scene perception may induce or even necessitate visual narrative-specific processes.

The development of SPECT has been carried out through tests of major assumptions of the framework. The BDF stories are an ideal stimulus set for these tests, because they are purely visual narratives that are easily manipulated (e.g., to require bridging inference generation). The BDF studies presented here tested how generalizable some important back-end mechanisms, which have been primarily studied in the context of textual narrative comprehension, are to the context of visual narrative comprehension. We found strong generalizability for the relationship between mapping and shifting mechanisms as a function of the degree of change in event indices, and likewise for the mapping mechanism sub-process of bridging inference generation. Nevertheless, each of these back-end mechanisms showed interesting differences in the context of visual narrative comprehension that could be related to specific front-end visual processes. For example, the front-end mechanism of broad information extraction, or scene gist, was proposed as the reason behind the difference in the relative importance of spatial changes for event segmentation between text and visual narratives. Likewise, we have suggested that pictorial versus textual medium-specific features place constraints on attentional selection that may explain differences in eye movements during bridging inference generation, while viewing visual versus textual narratives. Thus, the above work shows that the back-end event model mechanisms tested (i.e., laying the foundation, inference generation, shifting) are the same ones identified in research on text comprehension. However, due to the differences in the visual narrative stimuli compared to text, research on scene perception introduces important new considerations. This is precisely where SPECT offers unique contributions to further our understanding of comprehension processes in visual narratives, which have become the modal form of narrative consumed by much of the population. This work only scratches the surface of visual narrative comprehension, and there are many additional fundamental questions that SPECT poses (Loschky et al., "SPECT").

We also tested important hypotheses about the proposed bidirectional relationships between the front-end and back-end processes. An important future question is how exactly the front-end mechanism of information extraction influences the back-end mechanism of laying the foundation for a new event model. Initial investigations of this question (Larson; Larson and Loschky) showed that gist processing of different levels of scene categories (i.e., superordinate level indoor vs. outdoor, and basic level kitchen vs. office) and actions (basic level, cooking vs. washing dishes) occurred at different times scales, going from the coarse superordinate scene category level, within 150 ms processing time, to the fine basic level actions, which required two full eye fixations. Further research has shown that extracting the gist of a scene facilitates recognizing a person's action within that scene (Larson and Lee). These studies illustrate the kinds of studies motivated by hypotheses generated by SPECT. Many more such studies (Loschky et al., "SPECT"), and eventually computational models, will be needed to fully test the theory.

Acknowledgments

We would like to thank Adam M. Larson, Morton Gernsbacher, and G. A. Radvansky for helpful discussions of SPECT.

Works Cited

Baggett, Patricia. "Structurally Equivalent Stories in Movie and Text and the Effect of the Medium on Recall." *Journal of Verbal Learning & Verbal Behavior*, vol. 18, no. 3, 1979, pp. 333–356. ScienceDirect, doi:10.1016/S0022-5371(79)90191-9.

Cichy, Radoslaw Martin, Aditya Khosla, Dimitrios Pantazis, Antonio Torralba, and Aude Oliva. "Comparison of Deep Neural Networks to Spatio-Temporal Cortical Dynamics of Human Visual Object Recognition Reveals Hierarchical Correspondence." *Nature*, vol. 6, 2016. Scientific Reports, doi:10.1038/srep27755.

Clark, Herbert H. "Inferences in Comprehension." *Basic Processes in Reading: Perception and Comprehension*, edited by David LaBerge, and S. Jay Samuels, Erlbaum, 1977, pp. 243–263.

Cohn, Neil. *The Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images*. Bloomsbury, 2013.

Cohn, Neil, and Eva Wittenberg. "Action Starring Narratives and Events: Structure and Inference in Visual Narrative Comprehension." *Journal of Cognitive Psychology*, vol. 27, no. 7, 2015, pp. 812–828. Taylor & Francis, doi:10.1080/20445911.2015.1051535.

Eckstein, Miguel P., Barbara A. Drescher, and Steven S. Shimozaki. "Attentional Cues in Real Scenes, Saccadic Targeting, and Bayesian Priors." *Psychological Science*, vol. 17, no. 11, 2006, pp. 973–980. Sage, doi:10.1111/j.1467–9280.2006.01815.x.

Fei-Fei, Li, Asha Iyer, Christof Koch, and Pietro Perona. "What Do We Perceive in a Glance of a Real-World Scene?" *Journal of Vision*, vol. 7, no. 1, 2007, pp. 1–29. doi:10.1167/7.1.10.

Findlay, John, and Robin Walker. "A Model of Saccade Generation Based on Parallel Processing and Competitive Inhibition." *Behavioral and Brain Sciences*, vol. 22, no. 4, 1999, pp. 661–721.

Foulsham, Tom, Dean Wybrow, and Neil Cohn. "Reading without Words: Eye Movements in the Comprehension of Comic Strips." *Applied Cognitive Psychology*, vol. 30, no. 4. 2016, pp. 566–579. Wiley, doi:10.1002/acp.3229.

Gernsbacher, Morton Ann. *Memory for Surface Information in Non-Verbal Stories: Parallels and Insights to Language Processes.* Dissertation, U of Texas, 1983.

———. "Surface Information Loss in Comprehension." *Cognitive Psychology*, vol. 17, no. 3, 1985, pp. 324–363. ScienceDirect, doi:10.1016/0010-0285(85)90012-x.

———. *Language Comprehension as Structure Building*. Erlbaum, 1990.

Glanemann, Reinhild. *To See or Not to See -Action Scenes Out of the Corner of the Eye*. Dissertation, U of Münster, 2008.

Graesser, Arthur C., Murray Singer, and Tom Trabasso. "Constructing Inferences during Narrative Text Comprehension." *Psychological Review*, vol. 101, no. 3, 1994, pp. 371–395. PsycNET, doi:10.1037/0033–295X.101.3.371.

Haberlandt, Karl F., and Arthur C. Graesser. "Component Processes in Text Comprehension and Some of Their Interactions." *Journal of Experimental Psychology: General*, vol. 114, no. 3, 1985, pp. 357–374.

Haberlandt, Karl F., Claire Berian, and Jennifer Sandson. "The Episode Schema in Story Processing." *Journal of Verbal Learning and Verbal Behavior*, vol. 19, no. 6, 1980, pp. 635–650.

Hegdé, Jay. "Time Course of Visual Perception: Coarse-To-Fine Processing and Beyond." *Progress in Neurobiology*, vol. 84, no. 4, 2008, pp. 405–439. ScienceDirect, doi:10.1016/j.pneurobio.2007.09.001.

Henderson, John M., and Andrew Hollingworth. "High-Level Scene Perception." *Annual Review of Psychology*, vol. 50, 1999, pp. 243–271. Annual Reviews, doi:10.1146/annurev.psych.50.1.243.

Hollingworth, Andrew, and John M. Henderson. "Accurate Visual Memory for Previously Attended Objects in Natural Scenes." *Journal of Experimental Psychology: Human Perception & Performance*, vol. 28, no. 1, 2002, pp. 113–136.

Huff, Markus, Tino G. K. Meitz, and Frank Papenmeier. "Changes in Situation Models Modulate Processes of Event Perception in Audiovisual Narratives." *Journal of Experimental Psychology*, vol. 40, no. 5, 2014, pp. 1377–1388. PsycNET, doi:10.1037/a0036780.

Hutson, John P., Joseph P. Magliano, and Lester C. Loschky. "Understanding Moment-to-Moment Processing of Visual Narratives." *Cognitive Science*. Accepted pending revisions.

Itti, Laurent, and Christof Koch. "Computational Modeling of Visual Attention." *Nature Reviews Neuroscience*, vol. 2, no. 3, 2001, pp. 194–203. Nature, doi:10.1038/35058500.

Just, Marcel A., and Patricia A. Carpenter. "A Theory of Reading: From Eye Fixations to Comprehension." *Psychological Review*, vol. 87, no. 4, 1980, pp. 329–354. PsycNET, doi:10.1037/0033–295X.87.4.329.

Kurby, Christopher A., and Jeffrey M. Zacks. "Segmentation in the Perception and Memory of Events." *Trends in Cognitive Sciences*, vol. 12, no. 2, 2008, p. 72. ScienceDirect, doi:10.1016/j.tics.2007.11.004.

———. "Starting from Scratch and Building Brick by Brick in Comprehension." *Memory & Cognition*, vol. 40, no. 5, 2012, pp. 812–826. SpringerLink, doi:10.3758/s13421-011-0179-8.

Larson, Adam M. *Recognizing the Setting before Reporting the Action: Investigating How Visual Events Are Mentally Constructed from Scene Images*. Dissertation, Kansas State U, 2012.

Larson, Adam. M., and Lester C. Loschky. *From Scene Perception to Event Conception: How Scene Gist Informs Event Perception*. Submitted.

Larson, Adam M., and Melinda Lee. "When Does Scene Categorization Inform Action Recognition?" *Journal of Vision*, vol. 15, no. 12, 2015, p. 118. doi:10.1167/15.12.118.

Laubrock, Jochen, Anke Cajar, and Ralf Engbert. "Control of Fixation Duration during Scene Viewing by Interaction of Foveal and Peripheral Processing." *Journal of Vision*, vol. 13, no. 12, 2013, pp. 1–20. doi:10.1167/13.12.11.

Loschky, Lester C., and Adam M. Larson. "The Natural/Man-Made Distinction is Made Prior to Basic-Level Distinctions in Scene Gist Processing." *Visual Cognition,* vol. 18, no. 4, 2010, pp. 513–536. Taylor & Francis, doi:10.1080/13506280902937606.

Loschky, Lester C., Adam M. Larson, Tim J. Smith, and Joseph P. Magliano. The Scene Perception and Event Comprehension Theory (SPECT), Kansas State University. In preparation.

Loschky, Lester C., John P. Hutson, Joseph P. Magliano, Adam M. Larson, and Tim J. Smith. "Explaining the Film Comprehension/Attention Relationship

with the Scene Perception and Event Comprehension Theory (SPECT)." *SCSMI Conference*, 11–14 June 2014, Lancaster.
Loschky, Lester C., Joseph P. Magliano, and Tim J. Smith. "The Scene Perception and Event Comprehension Theory (SPECT) Applied to Visual Narratives." International Conference on Empirical Studies of Literature and Media*,* 6–9 July 2016, Chicago.
Magliano, Joseph P., Adam M. Larson, Karyn Higgs, and Lester C. Loschky. "The Relative Roles of Visuospatial and Linguistic Working Memory Systems in Generating Inferences during Visual Narrative Comprehension." *Memory & Cognition*, vol. 44, no. 2, 2016, pp. 207–219. SpringerLink, doi:10.3758/s13421-015-0558-7.
Magliano, Joseph P., and Jeffrey M. Zacks. "The Impact of Continuity Editing in Narrative Film on Event Segmentation." *Cognitive Science*, vol. 35, no. 8, 2011, pp. 1489–1517. Wiley, doi:10.1111/j.1551-6709.2011.01202.x.
Magliano, Joseph P., Kristopher Kopp, Karyn Higgs, and David N. Rapp. "Filling in the Gaps: Memory Implications for Inferring Missing Content in Graphic Narratives." *Discourse Processes*, vol. 54, no. 8, 2017, pp. 569–582. Taylor & Francis, doi:10.1080/0163853X.2015.1136870.
Magliano, Joseph P., Kristopher Kopp, M. Windy McNerney, Gabriel A. Radvansky, and Jeffrey M. Zacks. "Aging and Perceived Event Structure as a Function of Modality." *Aging, Neuropsychology, and Cognition*, vol. 19, no. 1–2, 2011, pp. 264–282. Taylor & Francis, doi:10.1080/13825585.2011.633159.
Magliano, Joseph P., Lester C. Loschky, James A. Clinton, and Adam M. Larson. "Is Reading the Same as Viewing? An Exploration of the Similarities and Differences between Processing Text- and Visually Based Narratives." *Unraveling the Behavioral, Neurobiological, and Genetic Components of Reading Comprehension*, edited by B. Miller, P. Brookes, et al., Paul H. Brookes Publishing: The Extraordinary Brain Series, 2013, pp. 78–90.
Mayer, Mercer. A Boy, a Dog, and a Frog. Dial Books, 1967.
———. Frog, Where Are You? Dial Books, 1969.
———. One Frog Too Many. Dial Books, 1975.
McNamara, Danielle S., and Joseph P. Magliano. "Toward a Comprehensive Model of Comprehension." *Psychology of Learning and Motivation*, edited by Brian H. Ross, Elsevier, 2009, pp. 297–384.
Mital, Parag K., Tim J. Smith, Robin Hill, and John M. Henderson. "Clustering of Gaze during Dynamic Scene Viewing is Predicted by Motion." *Cognitive Computation*, vol. 3. no. 1, 2010, pp. 5–24. SpringerLink, doi:10.1007/s12559-010-9074-z.
Moss, Jarrod, Christian D. Schunn, Walter Schneider, Danielle S. McNamara, and Kurt VanLehn. "The Neural Correlates of Strategic Reading Comprehension: Cognitive Control and Discourse Comprehension." *NeuroImage*, vol. 58, no. 2, 2011, pp. 675–686. ScienceDirect, doi:10.1016/j.neuroimage.2011.06.034.
Myers, Jerome L., Anne E. Cook, Gretchen Kambe, Robert A. Mason, and Edward J. O'Brien. "Semantic and Episodic Effects on Bridging Inferences." *Discourse Processes*, vol. 29, no. 3, 2000, pp. 179–199. Taylor & Francis, doi:10.1207/S15326950dp2903_1.

O'Brien, Edward J., Dolores M. Shank, Jerome L. Myers, and Keith Rayner "Elaborative Inferences during Reading: Do They Occur On-Line?" *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 14, no. 3, 1988, pp. 410–420. PsycNET, doi:10.1037/0278-7393.14.3.410.

Oliva, Aude. "Gist of a Scene." Neurobiology of Attention, edited by Laurent Itti, et al., Elsevier Academic P, 2005, pp. 251–256.

Pannasch, Sebastian, Jens R. Helmert, Katharina Roth, Ann-Katrin Herbold, and Henrik Walter. "Visual Fixation Durations and Saccade Amplitudes: Shifting Relationship in a Variety of Conditions." *Journal of Eye Movement Research*, vol. 2, no. 2, 2008, pp. 1–19. doi:10.16910/jemr.2.2.4.

Pertzov, Yoni, Galia Avidan, and Ehud Zohary. "Accumulation of Visual Information across Multiple Fixations." *Journal of Vision*, vol. 9, no. 10, 2009, pp. 1–12.

Ramkumar, Pavan, Bruce C. Hansen, Sebastian Pannasch, and Lester C. Loschky. "Visual Information Representation and Rapid-Scene Categorization Are Simultaneous across Cortex: An MEG Study." *NeuroImage*, vol. 134, 2016, pp. 295–304. ScienceDirect, doi:10.1016/j.neuroimage.2016.03.027.

Rayner, Keith. "Eye Movements in Reading and Information Processing: 20 Years of Research." *Psychological Bulletin*, vol. 124, no. 3, 1998, pp. 372–422. PsycNET, doi:10.1037//0033-2909.124.3.372.

Schmalhofer, Franz, and Doris Glavanov. "Three Components of Understanding of a Programmer's Manual: Verbatim, Propositional, and Situational Representations." *Journal of Memory & Language*, vol. 25, no. 3, 1986, pp. 279–294. ScienceDirect, doi:10.1016/0749-596X(86)90002-1.

Schyns, Philippe G., and Aude Oliva. "From Blobs to Boundary Edges: Evidence for Time- and Spatial-Scale-Dependent Scene Recognition." *Psychological Science*, vol. 5, no. 4, 1994, pp. 195–200. PsycNET, doi:10.1111/j.1467-9280.1994.tb00500.x.

Singer, Murray, and Michael Halldorson. "Constructing and Validating Motive Bridging Inferences." *Cognitive Psychology*, vol. 30, no. 1, 1996, pp. 1–38.

Smith, Marverick E., John P. Hutson, Joseph P. Magliano, and Lester C. Loschky. *Bridging the Gap in Coherence: Inference Generation and Event Segmentation Inform Event Model Construction Processes in Visual Narratives*. Kansas State University. In preparation.

———. *Laying the Foundation for Event Understanding in Picture Stories: Evidence from Eye Movements*. Kansas State University. In preparation.

Smith, Tim J. *An Attentional Theory of Continuity Editing*. Dissertation, U of Edinburgh, 2005.

———."The Attentional Theory of Cinematic Continuity." *Projections*, vol. 6, no. 1, 2012, pp. 1–27. Berghahn, doi:10.3167/proj.2012.060102.

Smith, Tim J., and Parag K. Mital. "Attentional Synchrony and the Influence of Viewing Task on Gaze Behaviour in Static and Dynamic Scenes." *Journal of Vision*, vol. 13, no. 8, 2013, pp. 1–24. doi:10.1167/13.8.16.

Swets, Benjamin, and Christopher A. Kurby. "Eye Movements Reveal the Influence of Event Structure on Reading Behavior." *Cognitive Science*, vol. 40, no. 2, 2016, pp. 466–480. Wiley, doi:10.1111/cogs.12240.

Torralba, Antonio, Aude Oliva, Monica S. Castelhano, and John M. Henderson. "Contextual Guidance of Eye Movements and Attention in Real-World Scenes:

The Role of Global Features in Object Search." *Psychological Review*, vol. 113, no. 4, 2006, pp. 766–786. PsycNET, doi:10.1037/0033–295X.113.4.766.

Wolfe, Jeremy M., Melissa L.-H. Võ, Karla K. Evans, and Michelle R. Greene. "Visual Search in Scenes Involves Selective and Nonselective Pathways." *Trends in Cognitive Sciences*, vol. 15, no. 2, 2011, pp. 77–84. PMC, doi:10.1016/j.tics.2010.12.001.

Zacks, Jeffrey M., Nicole K. Speer, and Jeremy R. Reynolds. "Segmentation in Reading and Film Comprehension." *Journal of Experimental Psychology: General*, vol. 138, no. 2, 2009, pp. 307–327. PsycNet, doi:10.1037/a0015305.

Zacks, Jeffrey M., Nicole K. Speer, Khena M. Swallow, Todd S. Braver, and Jeremy R. Reynolds. "Event Perception: A Mind-Brain Perspective." *Psychological Bulletin*, vol. 133, no. 2, 2007, pp. 273–293. PsycNET, doi:10.1037/0033–2909.133.2.273.

Zwaan, Rolf A., and Gabriel A. Radvansky. "Situation Models in Language Comprehension and Memory." *Psychological Bulletin*, vol. 123, no. 2, 1998, pp. 162–185. PsycNET, doi:10.1037/0033-2909.123.2.162.

Zwaan, Rolf A., Joseph P. Magliano, and Arthur C. Graesser. "Dimensions of Situation Model Construction in Narrative Comprehension." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 21, no. 2, 1995, pp. 386–397. PsycNET, doi:10.1037/0278-7393.21.2.386.