DISSOCIATING EYE-MOVEMENTS AND COMPREHENSION DURING FILM VIEWING

by

JOHN HUTSON

B.S., Knox College, 2010

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Psychological Sciences
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2016

Approved by:

Major Professor
Lester Loschky

Copyright

Abstract

Film is a ubiquitous medium. However, the process by which we comprehend film narratives is not well understood. Reading research has shown a strong connection between eye-movements and comprehension. In four experiments we tested whether the eye-movement and comprehension relationship held for films. This was done by manipulating viewer comprehension by starting participants at different points in a film, and then tracking their eyes. Overall, the manipulation created large differences in comprehension, but only found small difference in eye-movements. In a condition of the final experiment, a task manipulation was designed to prioritize different stimulus features. This task manipulation created large differences in eye-movements when compared to participants freely viewing the clip. These results indicate that with the implicit task of narrative comprehension, top-down comprehension processes have little effect on eye-movements. To allow for strong, volitional top-down control of eye-movements in film, task manipulations need to make features that are important to comprehension irrelevant to the task.

Table of Contents

List of Figures

Acknowledgements

I thank my advisor Dr. Lester Loschky for his guidance through each step of this project and continued support and dedication to advancing my development as a scholar. Many thanks also go to committee members Dr. Tim Smith (Birkbeck, University of London) and Dr. Don Saucier (Kansas State University) for their help through the course of completing this project. Dr. Smith provided important assistance with data analyses. Additionally, although not a member of the Thesis committee, Dr. Joe Magliano (Northern Illinois University) has been involved with the project from its inception, and has contributed to its development throughout. Thank you to Dr. Heather Bailey (Kansas State University) for her assistance collecting, scoring, and analyzing the working memory and event segmentation data. Finally, thank you to all the members of the Visual Cognition Lab at Kansas State University who have contributed to this work. Special thanks go to former lab graduate student and post-doctoral researcher Dr. Adam Larson, graduate students Ryan Ringer, Jared Peterson, and Maverick Smith, and all the lab research assistants who helped with data collection and discussion of this work.

Chapter 1 - Introduction

Watching highly produced dynamic scenes such as film, television, and online video is a ubiquitous activity around the world. People seem to comprehend these dynamic scenes without much effortful processing, but little is known about the processes involved in this comprehension (Smith, Levin, & Cutting, 2012). Furthermore, while watching a film, viewers typically move their eyes 2-5 times per second in order to extract information from it, and those eye-movements are most likely related to viewers' understanding of the film they are watching (Smith, 2013). Causally, this relationship can go in two directions: attention leading to comprehension differences, or what is being comprehended guiding attention. The primary concern of the current study is on the latter case. Namely, does a viewer's comprehension of a film influence their eye-movements while watching it? Although little previous research has addressed this question, there are two well-developed lines of previous research that are highly relevant: research on eye-movements and reading comprehension and research on eye-movements in static and dynamic scenes.

Top-down and bottom-up effects on eye-movements

At a broad level, comprehension processes for narrative content have been studied in the realm of language processing, typically in the context of reading (McNamara & Magliano, 2009, for review), and it has been shown that readers' eye-movements differ based on their comprehension of what they are reading (Rayner, 1998, for review). For example, during reading it has been shown that systematic regressive eye-movements are made when readers realize they have incorrectly interpreted something earlier in the story (Frazier & Rayner, 1982). Findings such as these are the basis of the *eye-mind hypothesis* (Just & Carpenter, 1980; Reichle, Pollatsek, Fisher, & Rayner, 1998; Reilly & Radach, 2006) that eye-movements are driven by online cognitive processes. From this, one would expect a connection between each movie viewer's comprehension and their eye-movements. More specifically, when movie viewers have different information incorporated into their mental models (Johnson-Laird, 1983) which they use to comprehend a narrative film, it seems reasonable that they would attend to different aspects of the film stimulus in order to update their mental model, namely the information that is present in their mental models will influence the information they attend to in the film.

Similar top-down effects on attention are found during scene viewing. When viewing static scenes, eye-movements can be affected by volitional top-down processes such as the goal or task of the viewer (Henderson, 2007; Henderson & Hollingworth, 1998, for review), and by more mandatory top-down processes such as looking at faces (Baluch & Itti, 2011; Birmingham, Bischof, & Kingstone, 2008). The same is true when watching video clips, in which the where and when of viewer attention on a screen is influenced by both volitional processes such as the goals of the viewer (Smith & Mital, 2013), and mandatory processes such as who is speaking (Coutrot & Guyader, 2014; Ho, Foulsham, & Kingstone, 2015; Vo, Smith, Mital, & Henderson, 2012). Alternatively, bottom-up features of scenes (i.e., features of the stimulus such as color, edges, and motion) are also known to have strong effects on visual attention (Itti, 2005; Mital, Smith, Hill, & Henderson, 2010), without affecting the interpretation of the scene (Latif, Gehmacher, Castelhano, & Munhall, 2014). In dynamic scenes the role of bottom-up features are thought to be very strong, such that when viewing highly produced dynamic scenes like a Hollywood film trailer, people tend to look in that same places *at the same time*, known as attentional synchrony. (Dorr, Martinetz, Gegenfurtner, & Barth, 2010; Hasson et al., 2008; Smith & Mital, 2013). This is very different from static scenes and natural dynamic scenes in which viewers tend to look at similar points of interest, but not at the same time (Mannan, Ruddock, & Wooding, 1997).

Film narrative is unique

There are differences between the linguistic and visual modalities of narrative representation that need to be accounted for when researching comprehension in visual narratives (Magliano, Loschky, Clinton, & Larson, 2013). For example, written text is composed of distinct words arranged in lines and paragraphs on a page, and readers typically fixate every content word (noun, verb, adjective, and adverb) in a line, progressing from left to right (in English). In contrast, films are composed of moving images within a frame, but there are no stated rules for how film viewers should watch them, though filmmakers follow numerous conventions in creating them (Smith, 2012). Also, film shots are typically viewed serially from beginning to end, unless a solitary film viewer uses a remote control with pause and rewind functions. This is in contrast with reading in which the reader controls their pace of reading and makes regressive eye-movements when they have difficulty understanding something.

Similarly, the highly produced nature of film contains several features that exert strong bottom-up control and increase attentional synchrony (Smith, 2013). The bottom-up features include motion (Mital et al., 2010), editing (Wang, Freeman, Merriam, Hasson, & Heeger, 2012), and lighting (Cutting, Brunick, DeLong, Iricinschi, & Candan, 2011; Smith & Mital, 2013). Additionally, highly produced dynamic scenes often compose scenes to include few points of interest, or are constructed such that the bottom-up features guide attention to a single point of interest (Cutting, 2015). Compared to highly produced film, the features of static text, static scenes, and natural dynamic scenes have relatively weak bottom-up features, and thus many studies have shown strong top-down effects on eye-movements when using such stimuli (Rayner, 1998; Smith & Mital, 2013; Yarbus, 1967). All of the above differences between films with reading and other types of scene viewing suggest that a simple analogy between how viewers process each is likely to be wrong.

Comprehension and eye-movements in film

The few studies that have tested top-down effects on eye-movements in film have what appear to be contradictory effects. Lahnakoski et al. (2014) found that giving viewers an explicit task to take a certain perspective (interior decorator or detective) can have a top-down effect on eye-movements. Alternatively, in an earlier study testing the same research question of how comprehension processes affect eye-movements, Loschky, Larson, Magliano, and Smith (2015) presented participants with a scene from the James Bond film *Moonraker* (Broccoli & Gilbert, 1979). They found that participants had large differences in comprehension, but there were relatively weak effects of comprehension on eye-movements. The eye-movement differences occurred during a single shot of the clip that was essentially a *static* image that allowed participant gaze to explore the image. In other words, the static nature of the scene may have allowed for eye-movement differences similar to those found in previous experiments using static scenes (DeAngelus & Pelz, 2009; Smith & Mital, 2013; Yarbus, 1967). Nonetheless, the overall lack of eye-movement differences were striking given the large effects typically found during static scene viewing, and the effects found for perspective taking (Lahnakoski et al., 2014) and location based viewing tasks (Smith & Mital, 2013). Smith & Mital (2013) and Taya, Windridge, & Osman (2012) give some converging evidence for a lack of top-down effects during conditions similar to free-viewing. This raises the critically important question addressed

in the current study, namely, why and when is there a general dissociation between eye-movements and film *comprehension*, which fails to support the eye-mind hypothesis in film viewing?

What may allow for top-down effects of attention in film?

To create a strong test of top-down effects in film, criteria were developed to choose a clip based on its bottom-up features and what it afforded in terms of top-down manipulations. First, the clip needed to lack specific bottom-up features that create attentional synchrony, which should enhance the opportunity for top-down processes to differentially guide viewers' eye-movements while watching the clip. Many film sequences show only a single primary object of interest in each shot, such as the James Bond *Moonraker* clip used in Loschky et al. (2015), which limits the opportunities for attention to be shifted to different screen locations. Conversely, some film segments contain many different things to look at, which should reduce the degree of attentional synchrony as different people may look at different things in the film frame. Likewise, each time there is a film cut (i.e., a switch between camera shots) there is a sudden decrease and then increase in attentional synchrony as viewers search for and then find the point of central interest in the new shot (Carmi & Itti, 2006; Mital et al., 2010; Wang et al., 2012). However, on rare occasions there are film sequences lacking any cuts for long periods of time (long-takes), thus removing the "resetting" happening after each cut. We chose one of the most famous long-takes in film history, the opening scene of the re-edited[1] version of Orson Welles' *Touch of Evil* (Welles, 1958). What makes this film opening so famous is that it is a very complex scene with many characters and points of visual interest to potentially look at that unfolds over three minutes. Although this clip has all the filmmaking mastery of a typical Hollywood style film, in comparison to the *Moonraker* clip (Broccoli & Gilbert, 1979) used in Loschky et al. (2015), it is much closer to a natural scene (i.e., no cuts, and many different things

[1] There are two major differences between the original and re-edited versions of *Touch of Evil*. First, the original used the opening scene for the credits of the movie. These were removed from the re-edited version so that all the visual information on the screen is directly related to the story being told. Second, the sound mix was redone to only include diegetic (i.e., meaningful real-world) sounds instead of having a film score playing over the entire scene.

to look at in the film frame), and thus it may allow for more influence of top-down processes involved in comprehension to differentially guide viewers' eye-movements.

To create top-down effects it may also be necessary to require the viewer to acquire information from different regions within the scene. For example, Taya and colleagues (2012) did not find differences in eye-movements when experts and novices watched a tennis match. One likely reason for this is that regardless of expertise, there was only one activity to watch—namely the ball and the player whose court the ball was in. Conversely, Smith and Mital (2013) asked participants to free view a dynamic scene or to identify the location it was filmed in. When identifying the location, participants had to focus on the background, while during free viewing participants could look at the people and actions occurring in the scene. For the current study, choosing a clip with multiple objects allows for a top-down manipulation that may require viewers to look in different places.

The narrative content of the opening shot of *Touch of Evil* allows for just such a manipulation of comprehension at the situation model level (Kintsch, 1998; van Dijk & Kintsch, 1983). This is a manipulation of viewer's mental representation based on the information available about the events in the film. The clip opens on a close-up of someone setting a time bomb (Figure 1). The time bomb is then placed into the trunk of a car, after which a couple unknowingly gets into the car and drive off, as the camera follows them. Importantly, after the bomb is put into the car, it is not seen again for the remainder of the clip. During viewing this creates a very suspenseful experience for the viewer as they wait for the time bomb to explode as they watch the events of the scene unfold. Knowledge of the bomb is what makes the clip so powerful. Without the bomb, it is just a mundane shot of people and cars on a street. Therefore, our comprehension manipulation plays on the power of the bomb to create suspense in what would otherwise be a mundane scene. Knowledge of the bomb thus creates a situation where viewers might pay more attention to the car, while those without knowledge of the bomb may be freer to explore other objects and characters in the scene.

To manipulate knowledge of the bomb, we used the jumped-in-the-middle paradigm developed by Loschky et al. (2015). This manipulation creates the common experience of coming into a television program or film part way through and then trying to comprehend what is happening. Specifically, there are two groups, Context and No-context, in which the Context group participants see the entire clip, while the No-context group miss a certain portion of the

clip at the beginning—in this case, the portion showing a time bomb being put into the car. Throughout the four experiments in this study, there were three starting points used. For all four experiments, the Context condition started at the beginning of the opening scene of *Touch of Evil* with a close up of the bomb that is then placed in the car. After the bomb is placed in the car and the villain runs away, a couple who are unaware of the bomb gets into the car and begins to drive down busy streets going in and out of the shot. Halfway through the clip a walking couple (Mr. and Mrs. Vargas) is introduced as new protagonists as the camera begins to follow them. Towards the end of the clip (Figure 1), the walking couple and the couple in the car reach a border checkpoint. After some time there, the walking couple passes through, with the clip ending as the walking couple are shown kissing in a close up. The No-context condition for Experiments 1 and 2 started just after the bomb is placed in the car, as the couple walked up to the car. Thus, the comprehension manipulation in Experiments 1 and 2 was knowledge of the bomb. In Experiments 3 and 4 the No-context group started watching the film while the walking couple was on the screen but the car was off-screen. With this alternative No-context manipulation there were two main differences between the context groups: knowledge of the bomb and the perceived protagonists of the clip. The Context group may have perceived the couple in the car, and potentially the walking couple, as protagonists, while the No-context group were likely to perceive *only* the walking couple as the protagonists. Along with testing whether weaker bottom-up features would allow for a stronger effect of top-down attention on eye-movements, having two No-context conditions that manipulated different aspects of viewers' situation models (hereafter referred to more generically as *mental models*) allowed us to test different information sources used for comprehension. Finally, an additional condition in Experiment 4 tested the effect of task on eye-movements similar to Smith & Mital (2013) and Lahnakoski et al. (2014). The task manipulation made comprehension for the film narrative irrelevant, which allowed for comparisons of effects due to comprehension versus task.

Figure 1. Frames illustrating important shots in the 3 minute 12 second clip from the film *Touch of Evil* (Welles, 1958).

The manipulation of context (i.e., whether participants saw the bomb) should lead to different mental models for the target segment of film (e.g., Bransford & Johnson, 1972). That is, in the Context condition (i.e., viewers who saw the bomb placed in the car), a token for the car should be represented in participants' mental models for the film segment. That token should be reactivated every time the car is in the frame, which should then lead to the reactivation of the memory representation of the bomb and the life-threatening events it could cause for characters near it (e.g., Myers, & O'Brien, 1998). Conversely, in the No-context condition (i.e., viewers who did not see the bomb put in the car), the car had no particularly salient causal connections in the unfolding narrative, other than as a means of transportation for the couple in it. Thus, it should simply be a part of the backgrounded events and should be relatively weakly represented in the mental model. It is therefore reasonable to hypothesize that there would be a greater likelihood of viewers fixating on the car in the Context condition than in the No-context condition. Alternatively, one might hypothesize that consistent with the results of Loschky et al. (2015), the attentional synchrony created by the bottom-up features of this highly produced film

might wash out any differences in eye-movements that would be expected to occur based on differences in viewers' mental models. However, in comparison to the James Bond *Moonraker* clip used in the Loschky et al. (2015) study, the comparatively weaker bottom-up features of the *Touch of Evil* clip used should theoretically give the top-down comprehension processes a greater chance to guide attention. Based on these arguments, the two competing hypotheses for this study were:

> *Mental Model Hypothesis*: Top-down processes (involved in comprehension) will guide viewer attention creating low attentional synchrony between Context and No-context conditions
>
> *Tyranny of Film Hypothesis*: Bottom-up features will guide attention and wash out any comprehension-based differences between groups

Support for the mental model hypothesis would involve the Context and No-context groups producing different eye-movement patterns while viewing the clip. Specifically, viewers in the Context condition with knowledge of the bomb would be expected to maintain that information in their mental model throughout the film clip, and therefore spend more time attending to bomb-relevant information in the scene, such as the car (and its trunk) containing the bomb. Conversely, in the No-context condition, viewers would not be expected to give as much attention to these bomb-relevant elements. Conversely, support for the tyranny of film hypothesis would be evidenced by both groups of viewers showing similar eye-movement patterns while watching the clip.

Study overview

This study adopted an experimental case study approach, in that it used a single film clip of the opening scene of *Touch of Evil* (Welles, 1958), and then experimentally manipulated viewers' comprehension of that clip in a number of experiments. Similar experimental case study approaches have been used in a number of psychological studies that involved highly ecologically valid stimuli for which it was difficult to produce multiple instantiations (Bransford & Johnson, 1972; Brewer & Treyens, 1981; Loschky et al., 2015; Simons & Chabris, 1999;

Simons & Levin, 1998; Smith, Lamont, & Henderson, 2012). Two sets of experiments were run, each starting with an experiment testing participant comprehension, and then using the comprehension results to support an eye-movement experiment.

In the first set of experiments (Experiments 1 and 2), Experiment 1 measured comprehension by asking participants to generate a predictive inference of what would happen next at the end of the video clip. It was expected that participants in the Context condition who knew about the bomb in the car would be more likely to infer an explosion (or related event) than participants in the No-context condition who did not know about the bomb. This first experiment was important for two reasons. First, it was used to establish that the jumped-in-the-middle paradigm produced a strong effect on comprehension. Second, it manipulated the presence of audio to test whether it would have an effect on inference generation. Experiment 2 then introduced eye-tracking to test the effects of such comprehension differences on eye-movements, but was otherwise procedurally the same as Experiment 1. Furthermore, Experiment 1 included audio and no-audio conditions, to test for the importance of sound in heightening viewer suspense in the clip (Tully, 1999).

In the second set of experiments (Experiments 3 and 4), the No-context condition was changed to manipulate not only knowledge of the bomb, but also the perceived protagonist(s) of the clip (see figure 1). In Experiment 3 using this new No-context condition, in addition to measuring comprehension at the end of the clip with the predictive inference measure, we also measured working memory to test whether participants higher in working memory were more likely to make the inference about the bomb in the Context condition. This was of interest based on studies of reading, in which individuals with higher working memory tend to have better comprehension (Allbritton, 2004; Calvo, 2001, 2005; Daneman & Carpenter, 1980; Daneman & Merikle, 1996; Just & Carpenter, 1992; Linderholm, 2002; Rai, Loschky, & Harris, 2014; Rai, Loschky, Harris, Peck, & Cook, 2011; St George, Mannes, & Hoffman, 1997). As an on-line measure of comprehension participants also completed an event segmentation task (Newtson, 1973). Event segmentation has been shown to be sensitive to viewers' perceived event structure of both natural and narrative films. Experiment 4 was split into sub-experiments 4a and 4b. Experiment 4a measured viewers' eye-movements during viewing in the Context and new No-context conditions. Experiment 4b added a condition that instructed participants to draw a map of the scene depicted. This tested the effect of a cognitive task manipulation at odds with

comprehending the narrative, similar to the task manipulations in Lahnakoski et al. (2014) and Smith & Mital (2013). The control condition helps to place the effect of comprehension on eye-movements in the larger theoretical context of top-down task effects on eye-movements.

Chapter 2 - Experiment 1: Predictive Inference Generation

Method

Participants

There were 94 participants included in data analyses (54 females; mean age = 18.6 years; *SD* = 1.3). A total of 116 participants started the experiment, but 22 did not give an inference at the end of the experiment[2]. All participants were pseudo-randomly assigned to one of four viewing conditions of the opening scene of *Touch of Evil*. The participants included in data analysis had a fairly equal representation in each condition (Context + Audio, n = 21; Context + No-Audio, n = 29, No-context + Audio, n = 24, No-context + No-Audio, n = 20)[3]. Participants were Kansas State University undergraduate students participating in the study for course research credit. Application to the University Institutional Review Board determined this and all following experiments in the study posed minimal risk to the participants (i.e., exempt under the criteria set forth in the Federal Policy for the Protection of Human Subjects), and informed consent was determined to be unnecessary.

Stimuli

Two clips from the opening scene of Orson Welles *Touch of Evil* were used (Welles, 1958). The Context version shows a bomb being placed in a car trunk at the beginning, and runs for 3:12. The No-context version omits the first 18 seconds when the bomb is placed in the car, and runs for 2:54. Both clips end with a close up of the walking couple kissing. The Context and No-context conditions were presented in both Audio and No-audio conditions. The presence of audio was manipulated because one of the key purposes of the audio track was to maintain the suspense of the bomb. To do this the car that had the bomb in it also had its radio playing. Throughout the clip as the car moves towards and away from the camera, the sound of the radio

[2] It is unclear why there was a relatively high dropout rate for this online experiment. However, the dropout rate did not appear to have affected the results, because the inference results from this experiment were replicated in the following in person experiments.

[3] The unequal number of participants in each condition occurred because not all participants completed the online study, and those participants were not included in any analyses. In total 116 participants began the study, with 22 not finishing it.

gets accordingly louder and softer. No mention of the bomb is made in the dialogue except for a comment referring to "hearing a ticking noise" made by the female passenger of the car near the very end of the clip. Adobe Premiere Pro was used to edit the clips into the two conditions at a frame rate of 30 frames per second (fps), and a resolution of 1080 x 720 pixels. As this was an online study, participants viewed the video on their personal computer at an unspecified viewing distance.

Procedure

In all four conditions the experimental procedure was the same. Participants first followed a link to the online study. Next, they were told they would see a short video clip. Participants were then randomly assigned to one of four Context x Audio conditions.

After viewing the video clip, participants were presented with a series of written questions. To check whether participants in the Context condition maintained the bomb in their mental model, the first question was, "What do you think will happen next?" and they were prompted to type their written response in a text box. The next two questions were to ensure participants had not seen the clip before. The second question was, "Have you seen this film before?" and they pressed a "Yes" or "No" button in response. If participants answered "Yes," they were presented with the third question, "What is the name of this film?" to which they were prompted to provide their written answer in a text box. No participants reported seeing the film before.

Data analysis

To identify whether participants' predictive inferences at the end of the clip were influenced by having the bomb in their mental model, we had two research assistants code each inference, with coders blind to the condition from which each response was taken. Coding was conducted with all 116 participants that began the online study. If no response was given, it was coded as '0' *and excluded from further analysis*; responses having no mention of the bomb or an explosion were coded as '1'; unclear responses were coded as '2'; responses making clear reference to the bomb or an explosion were coded as '3'. The coders had a relatively high level of inter-rater reliability, producing Cohen's *Kappa* = .892, $p < .001$. Discrepancies between the

two coders were resolved through discussion. Unclear responses (coded as '2') were resolved as relating to the bomb or not, resulting in a final dichotomous coding indicating whether participant predictions mentioned the bomb '1', or did not mention the bomb '0'.

Results & Discussion

Once coding was completed, Chi Square tests were used to test whether there were differences in the frequencies of making a prediction about the bomb, between the Context or No-context conditions and in the Audio or No-Audio conditions. As shown in Figure 2, unsurprisingly, there were large differences in whether participants made a bomb-relevant predictive inference based on their viewing context (X^2 (1, N = 94) = 28.517, p < .001; *Eta* = .551), with participants who saw the bomb at the beginning being much more likely (56%) to make an inference about it than those who did not see it (4.5%) [Figure 2]. Interestingly, within the Context condition, those in the Audio condition were much more likely to draw bomb-related inferences (76%) than those in the No-audio condition (41%)(X^2 (1, N = 50) = 5.99, p = .014; Eta = .346). Nevertheless, there was no significant difference between audio conditions when collapsed across context (X2 (1, N = 94) = 2.597, p = .107; Eta = .166), indicating that drawing the inference was primarily dependent on having prior knowledge of the bomb. For this reason it was quite surprising to find that two participants in the No-context +Audio condition did make a predictive inference about the bomb exploding, since they had not seen the bomb. These participants did not indicate having seen the film before. However, a possible explanation is that they were able to make the inference due to a character in the car (which contained the time bomb) mentioning "hearing a ticking noise" towards the end of the clip. This would also help explain why in the Context condition, the Audio group outperformed the No-audio group. Specifically, perhaps some viewers in the Context + Audio condition forgot about the bomb but were reminded of it near the end of the film clip when they heard the "ticking" comment. There was even one participant in the No-context + Audio condition who made the inference, which is more evidence that the "ticking" comment helped participants make the inference about the bomb. Conversely, if there were an equal proportion of participants in the Context + No-audio condition who forgot about the bomb, they would not have been reminded of the bomb, because they could not hear the "ticking" comment. If so, then the difference in performance between the Context + Audio versus the Context + No-audio groups (76% - 41% = 35%) would represent the

proportion of Context condition participants who forgot about the bomb. We later tested this hypothesis in Experiment 3. In addition, a possible future experiment to test whether the "ticking" comment was driving the Audio condition advantage could simply remove the comment from the audio condition.

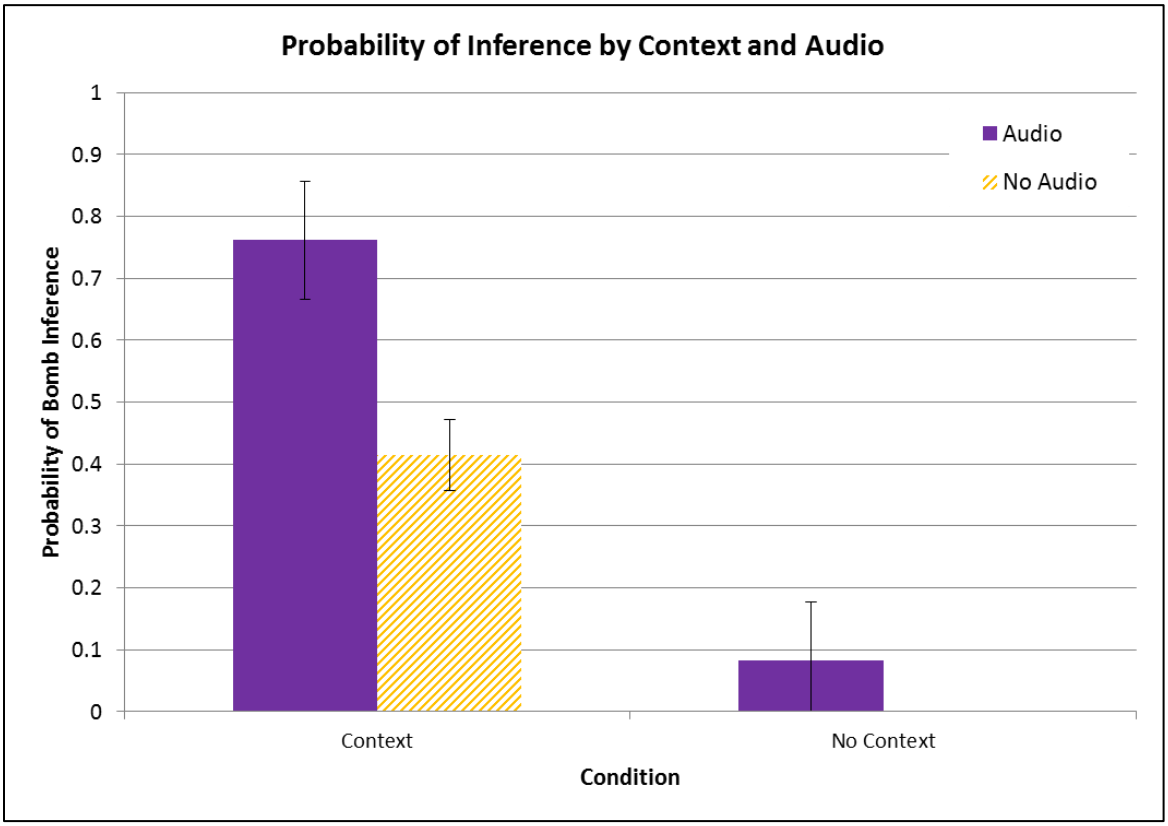


Figure 2. Probability of making inference about the bomb. Context condition is on the left, and No-context is on the right. Audio condition is in purple, and No-Audio is in striped yellow.

The strong effect of context (knowledge of the bomb) on making bomb-relevant inferences is not surprising. However, it is important for the study, because it establishes that the context manipulation produces clear differences in the information participants have in their mental models as they view the film clip. With participants showing this difference in inference generation and mental model construction, there is support for running the eye-tracking experiment to test whether differences in comprehension will produce differences in eye-movements. There are a number of eye-movement differences that could be expected due to the comprehension differences. First, participants in the Context condition that are following the

narrative should experience some suspense during viewing, which has been shown to narrow attentional focus during film viewing (Bezdek et al., 2015). This narrowing of attention could create even higher attentional synchrony among the Context condition participants. Second, related to the higher attentional synchrony, participants in the Context condition may be expected to be more likely to guide their attention to the car with the bomb in it, because this is the most important causal event index of the narrative event model for comprehension of the clip. If comprehension is guiding attention, No-context participants would be expected to explore the narrative space more, showing less attentional synchrony and fewer looks at the car since they do not know it contains the bomb.

Chapter 3 - Experiment 2: Context and Eye-movements

Method

Experiment 2 tested the effect of the comprehension differences produced in Experiment 1 on viewers' eye-movements while watching the film clip. Because differences in comprehension due to the Context manipulation were greatest in the Audio condition in Experiment 1 (see Figure 1), only the Audio condition was included in Experiment 2.

Participants

There were 84 participants (61 females; mean age = 18.6; *SD* = 1.4) who were pseudo-randomly assigned to one of two viewing conditions for the opening scene of *Touch of Evil* (Context: n = 42; No-context: n = 42). Participants were undergraduate students at Kansas State University who participated for course research credit.

Stimuli

The stimuli were identical to those in Experiment 1 in the Audio condition. The video clips were shown on a 17" ViewSonic Graphics Series CRT monitor (Model G90fb). A chin and forehead rest set a fixed viewing distance of 60.96 cm. The screen subtended 21.42° x 16.10° of visual angle.

Procedure

All participants were told that they would be shown a video clip while their eyes were tracked. Eye tracking was done using an EyeLink1000 eye tracker, which samples eye position 1000 times per second (1000 Hz) with an average spatial accuracy of 0.5° of visual angle, and a maximum error of 1°. Participants went through a nine point calibration routine, after which the experiment began. An eye-movement trigger was used to ensure that the video started at the beginning of a fixation. To start a trial, while the participant was looking at the central fixation point, they pressed a button which moved the fixation point 13.65° to right of center. Once the participant fixated the new point, it moved back to the center. During the saccade (velocity > 30°/sec) back to the center, the video began to play. In this way, any saccadic inhibition (which

increases the current fixation duration), caused by the motion transient due to the sudden onset of the video clip, was masked by the viewer's own eye-movement (Reingold & Stampe, 2000, 2002). Participants then watched the video, uninterrupted, until the moment when the couple kisses (3:12 into the Context condition and 2:54 into the No-context condition). At the end of the video all participants were asked, "What will happen next?" and responses were collected using the computer keyboard. The next question asked was, "Have you seen this movie before?" The keyboard was used to indicate "Yes" or "No." If a participant responded "Yes" they were asked the follow up question, "What was the name of the movie?" No participants indicated having seen the movie before.

Data analysis

In this experiment, the focus of data analysis was on participant condition rather than whether a bomb-relevant inference was made, for reasons explained after the inference results are reported. Inferences were coded by the same two research assistants as in Experiment 1 using similar procedures. For this experiment, the coding of the inference was dichotomous from the beginning (1 = participant mentioned something related to the bomb, 0 = the participant did not). Again, the coders had a high level of inter-rater reliability, producing Cohen's *Kappa* = .954, $p < .001$. Any remaining discrepancies between the two coders were resolved through discussion. After coding, the four participant groups were Context + Inference (n = 33), Context + No-inference (n = 9), No-context + Inference (n = 1), and No-context + No-inference (n = 41).

Results

Predictive inference

A chi-square test was used to identify if there was a difference between the Context and No-context conditions. As in Experiment 1, we found the expected large difference between context conditions, with 80% of participants in the Context condition making a bomb-relevant inference compared to only one participant from the No-context condition doing so (probably due to the "ticking" comment as in Experiment 1), X^2 (1, N = 85) = 51.59, $p < .001$.

These results indicate that participants had similar comprehension of the clips to participants in Experiment 1. In Experiment 1, we also pointed out that those in the Context +

Audio condition were twice as likely as those in the Context + No-Audio condition to make the inference, which could be the result if half of the participants in the Context condition forgot about the bomb until being reminded by hearing the "ticking comment" near the end of the clip. If this reasoning is correct, then it calls into question comparing the eye-movements of those who did versus did not make the inference in the Context condition in the current experiment, because all participants had audio. Specifically, for participants in the Context condition who made the inference, it is unknown whether their inference was based on their mental model having included the bomb from the beginning of the clip to the end, or instead based on suddenly remembering the bomb after hearing the "ticking noise" comment near the end of the clip. We directly addressed this issue in Experiment 3, but for the current experiment eye-movement analyses focus on participant condition, Context versus No-context, rather than inference. For comparison, analyses are included that removed participants in the Context condition who did not make the inference about the bomb.

Eye-movements

Fixation durations and saccade lengths

All eye-movement data were first cleaned by removing the longest and shortest 1 percent of fixation durations and saccade lengths for each participant. We then compared the mean fixation durations and saccade lengths between the Context and No-context groups for the shared viewing period. There were no significant differences in fixation duration between the 2 conditions. In the Context condition the average fixation duration was 392 milliseconds (SD = 65), and it was 386 (SD = 63) in the No-context condition (t (82) = .632, p = .529). The relationship held when only participants in the Context condition who made the inference were compared to the No-context condition (t (73) = .491, p = .625). The average saccade length for the Context group was 4.84° of visual angle (SD = .63°), and it was 4.63° (SD = .66°) for the No-context group (t (82) = 1.848, p = .068, d = .41). When only Context condition participants who made the inference were included the effect of condition and inference on saccades lengths was significant (t (73) = 2.089, p = .040, d = .489). Thus, participants in the context condition who made the inference had longer saccade lengths (4.89° of visual angle; SD = .58°) compared to the No-Context condition (4.63°; SD = .66°).

Longer saccade lengths usually show greater exploration of a scene (Pannasch, Helmert, Roth, Herbold, & Walter, 2008), which makes it surprising the Context group that made the inference would explore more. They have the best understanding of the narrative presented, and many of them maintained the bomb in their mental model throughout the clip. This should create suspense that would guide their eye-movements towards the car with the bomb, which should result in shorter saccade lengths. However, there is the possibility that due to their relatively good comprehension for the narrative Context participants were under less cognitive load to maintain the narrative, which gave them the opportunity to explore the screen more. This may be similar to a person watching a film for the second or third time, and noticing things they hadn't in previous viewings because they don't have to follow the narrative as closely.

Attentional synchrony

Data pre-processing. To calculate attentional synchrony the raw eye-tracking data was down sampled to 33Hz (from 1000 Hz) to express eye fixation X/Y coordinates per video frame, and exclude saccades and blinks. Gaze was then visualized using CARPE (Computational and Algorithmic Representation and Processing of Eye-movements) software developed by Mital et al. (2010). To identify whether the distribution of gaze differs between the two viewing conditions frame-by-frame we calculated the metric gaze similarity (Loschky et al., 2015; Mital et al., 2010), an adaptation of the Normalized Scanpath Saliency (NSS) first proposed by (Peters, Iyer, Itti, & Koch, 2005) and extended to video by Dorr et al. (2010) (for details of the method and equations, see Dorr et al. (2010)). We modify the NSS method in two critical ways. First, to calculate inter-observer similarity within the reference condition (in this case, Context), a probability map is created by plotting 2D circular Gaussians (1.2° SD % the fovea) around the gaze locations within a specific time window (225 ms; roughly equivalent to an average fixation) for all but one participant within the Context condition. These Gaussians are summed and normalized relative to the mean and SD of these values across the entire Context condition, in order to see how the similarity fluctuates over time (z-score similarity = (raw values– mean) / SD). The gaze location of the remaining participant is then sampled from this distribution (i.e., a Z-score is calculated for this participant) to identify how their gaze fits within the distribution at that moment. This leave-one-out procedure is repeated for all participants within the Context condition until each participant has a z-scored value (referred to as gaze similarity here). These

values express both 1) how each individual gaze location fits within the group at that moment and 2) how the average gaze similarity across all participants at that moment differs from other times in the video: A z-score close to zero indicates average synchrony, negative values indicate less synchrony than the mean (i.e., more variance), and positive values indicate more synchrony.

Second, the method is extended to allow gaze from different viewing conditions (e.g., No-context) to be sampled from a reference distribution (Context). For each gaze point in the No-context condition, the probability that it belongs to the Context condition's distribution is identified by sampling the value at that location from the Context's probability distribution (this time leave-one-out is not used as the gaze does not belong to the same distribution so cannot be sampled twice). The resulting raw NSS values for No-context are then normalized to the reference condition. Importantly, if the two distributions are identical, the average z-scored similarity for both distributions will fluctuate together, expressing more (positive z-score) or less (negative z-score) attentional synchrony over time (see Figs 3). However, as the similarity score is derived from the reference distribution if the two distributions differ significantly we cannot know if this is because the comparison distribution has more or less attentional synchrony than the reference distribution. We can only say that the comparison distribution differs more from the reference distribution than the reference distribution differs from itself. For example, both gaze distributions could be tightly clustered but in different, non-overlapping parts of the screen or the comparison distribution could be more spread out and only partly overlap with the tight reference distribution. Both situations would result in significant differences between the distributions.

Additionally, a shuffled baseline was created to simulate random eye-movements during the clip. The shuffled baseline started with the Context group. Because the attentional synchrony values are calculated based on each participant's gaze location on each film frame, the order of frames for each participant (and thus the order of their eye-movements across frames) was shuffled. In other words, for the first frame of the film, instead of having each participants first fixation the shuffled baseline may have one participants first fixation, a second participants 356^{th} fixation, and another's 22^{nd}, etc. This new gaze distribution was then compared to the reference distribution (i.e. Context) in the same method described above. This created a chance baseline of gaze similarity specific to visible contents the film clip.

Attentional synchrony results. The first gaze similarity analysis compared the Context and No-context conditions across the entirety of the film clip overlapping across conditions (2 minutes and 54 seconds of film). The upper section of Figure 3 illustrates the results of this comparison. The lower section of Figure 3 shows fixation heat maps for three frames from the film clip that illustrate both high and low levels of gaze similarity in both the Context and No-context conditions.

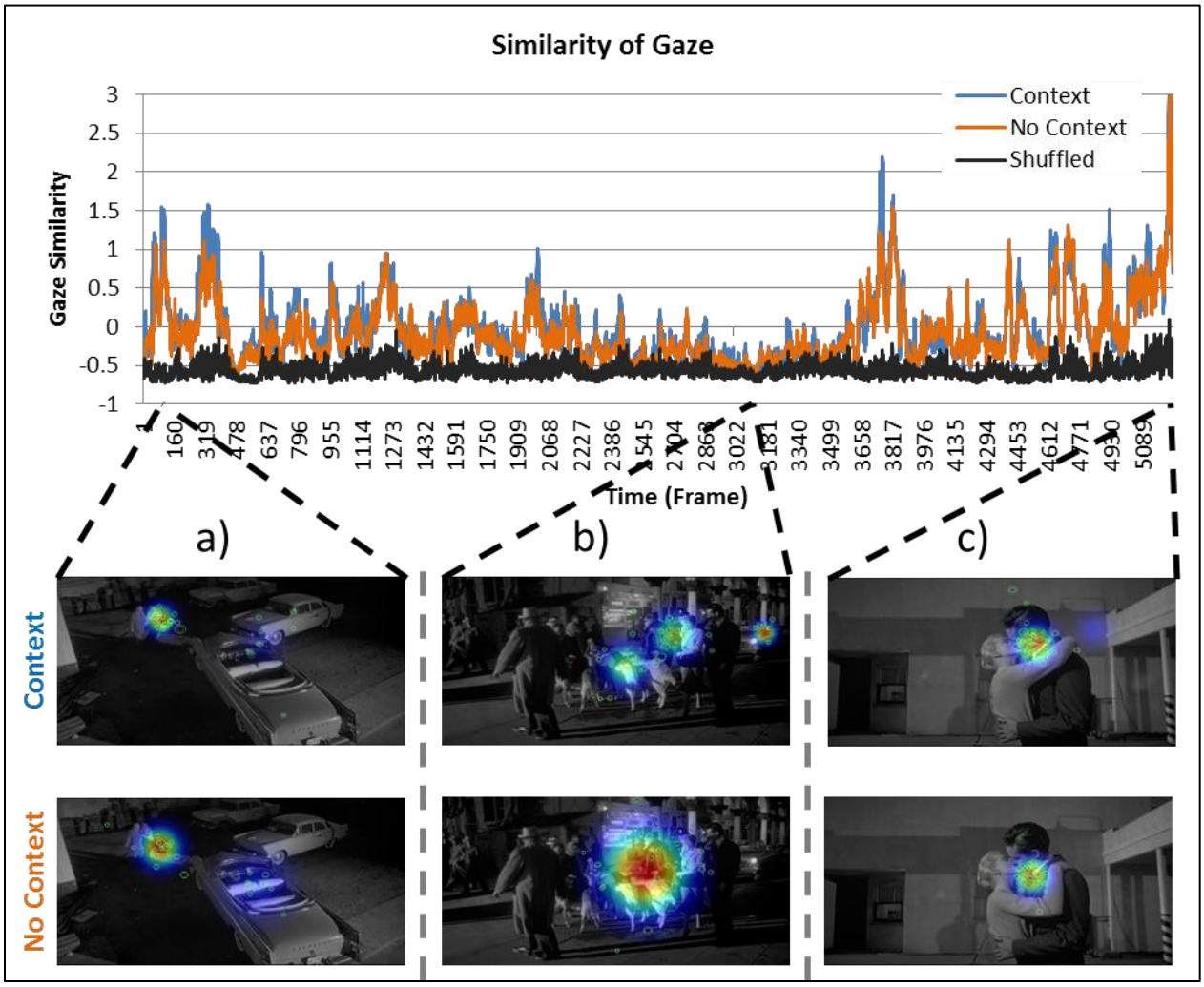


Figure 3. Top: Similarity of gaze by context condition across the shared viewing period of the clip. Gaze similarity is expressed as a z-score probability relative to the Context condition. (Context [Blue], No-context [Orange], and Shuffled Baseline [Black]). Large values indicate greater attentional synchrony. Bottom: Three of the peaks in gaze similarity are illustrated by image frames with superimposed heat maps of participant gaze location. The frames show the gaze heat maps at the points indicated on the gaze similarity figure for both the context and No-

context conditions. Frames a) and c) show high gaze similarity, while frame b) shows low gaze similarity. Note that Frame c) was the single highest level of attentional synchrony in the entire film clip.

Qualitatively, looking at this figure one can see that gaze similarity scores for the Context and No-context groups generally vary together, indicating that regardless of context condition, viewers had the same patterns of attentional synchrony. A *t*-test of mean gaze similarity by group supported this qualitative assessment (t (80) = 1.081, p = .283; d = 0.241), indicating that knowledge of the bomb did not have an effect on overall viewer attentional synchrony. The results are similar with participants that did not make the inference in the Context condition are removed from the analysis (t (71) = .592, p = .556). Next, the shuffled baseline is included for comparison. As shown in figure 2, when the experimental groups' gaze similarity is above the shuffled baseline, it indicates that the film is guiding eye-movements. With the shuffled baseline included in an ANOVA, it is statistically significant (F (2, 122) = 73.727, p < .001, η_{p2} = .551). Bonferroni corrected pairwise comparisons indicated that the shuffled baseline chance level of attentional synchrony (M = -.561, SD = .034) was significantly less than both the Context (M = -.001, SD = .267) and No-context (M = -.067, SD = .290) conditions.

Region of interest

Data pre-processing. Dynamic regions of interest were created for the clip to test whether either condition looked more at the car with the bomb in it. To create the dynamic region of interest for the car, we used Gazeatron (Smith & Mital, 2013; Vo et al., 2012) to identify the rectangular X and Y pixel coordinates for the car on the screen for each frame (at 30 frames/sec). These pixel coordinates were then exported and combined with the raw fixation report from EyeLink DataViewer (SR Research). This was used to calculate the cumulative dwell time and mean number of fixations for each participant in the car region of interest.

Region of interest results. While gaze similarity is a metric that indicates the co-occurrence of eye-movements in space and time, it does not indicate the features of a scene that are being attended to. The region of interest analysis remedies this by indicating how much a specific object in a scene, here the car with the bomb in it, is attended to. The Mental Model

Hypothesis predicted that the car with the bomb would be of greater importance to participants in the Context condition, because they are aware of the potential destructive causal effects the car could have on nearby persons, places, and things.

As with the gaze similarity analysis, fixations on the car by viewers in the two context conditions were compared for the shared viewing time from the start time of the No-context condition when both conditions were seeing the exact same information. The region of interest was used to calculate the mean number of fixations when the car was present on the screen within 30 frame (1 sec) time bins. The upper portion of Figure 4 illustrates the results of this comparison. The lower portion of Figure 4 shows fixation heat maps on two frames from the clip that illustrate high versus low rates of fixating the car region of interest in the Context and No-context conditions.

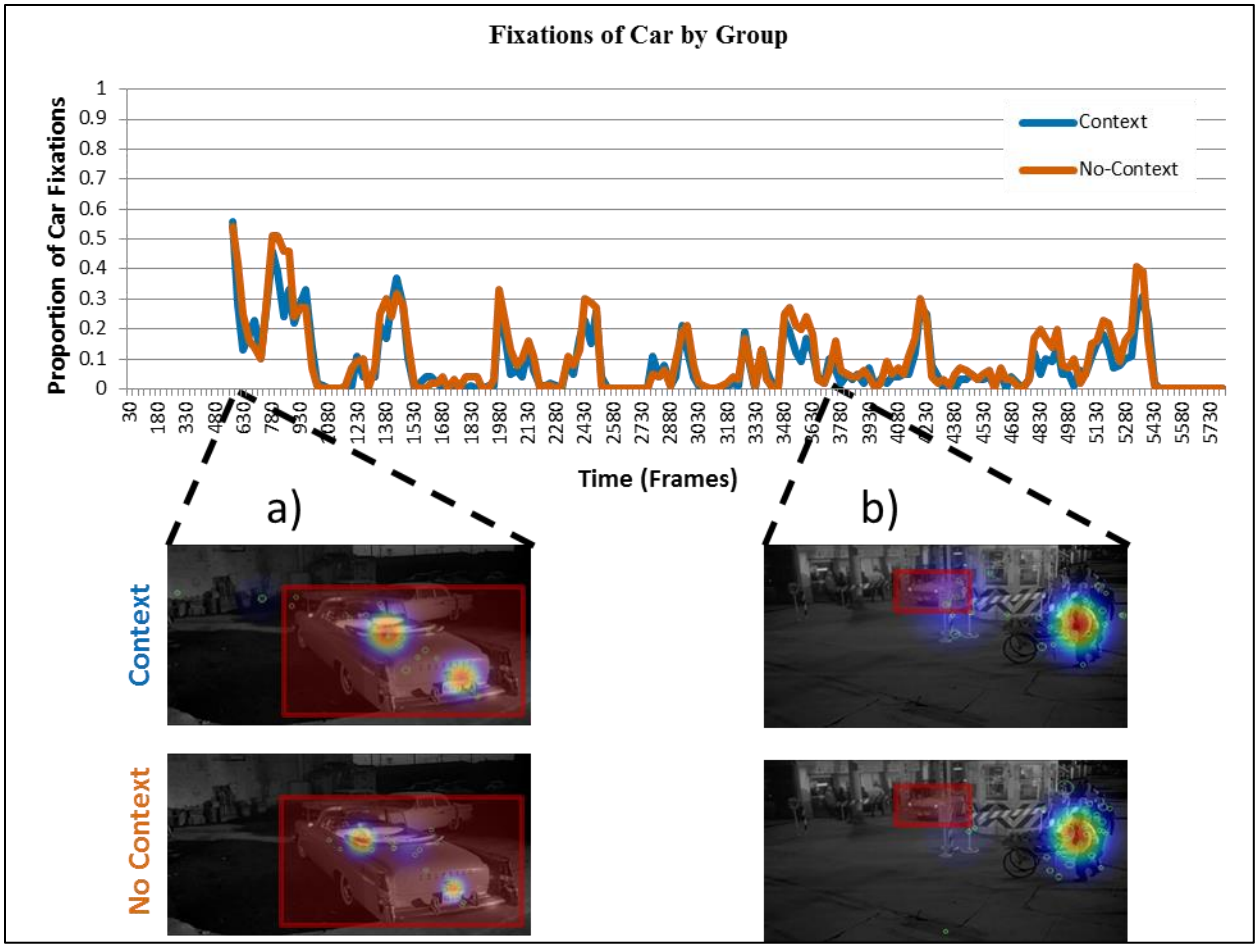


Figure 4. Top: Proportion of participants fixating the car by context condition throughout the film clip (3 second bins). The higher the value the more participants looking at the car. Bottom: Film stills show the region of interest for the frame, with fixation and heat maps superimposed. Still set a) shows a time point when the car was fixated by a majority of participants, and b) shows when the car was minimally fixated.[4] Note that frame a) shows the single highest proportion of fixating the car region of interest, at the start of the common viewing period across both context conditions.

As with the gaze similarity analyses, the lines for the two context conditions are mostly on top of one another, indicating that regardless of the context condition, participants seem to have fixated the car at the same time points throughout the clip. A *t*-test comparing the proportion of fixations on the car for each condition was not statistically significant (t (82) = 1.73, p = .087; *Cohen's d* = 0.382), but was trending in that direction. The trend, however, was for the No-

[4] Any apparent discrepancies between the proportion indicated and the number of fixations shown in the region of interest in the stills occur due to the binning across 30 frames used to create the data graph.

context group to have a higher proportion of fixations (M = .098, SD = .045) on the car than the Context group (M = .082, SD = .036), which is in the opposite direction of what was predicted. The result was the same when only those members of the context group who made the inference were included (t (73) = 1.434, p = .156; d = 0.316). Overall, this indicates that the viewers without knowledge of the bomb fixated the car at a similar rate as those with knowledge of it. A more targeted analysis was performed next to test whether there may be specific points where there are differences in the proportion of car fixations between the groups.

Latency to fixate the car after disocclusion

Both the gaze similarity and region of interest analysis compared eye-movement variables across the entire video clip. However, we felt those analyses may not have been sufficiently sensitive to find more subtle differences in eye-movements as a function of comprehension differences. Thus, the next analysis was designed to identify potential differences at specific time points where the comprehension literature might predict differences. Interestingly, in the film clip, the car with the bomb comes in and out of view at various points, creating suspense for the viewer. Thus, we tested whether, after the car had been out of view, when it came back into view (i.e., was disoccluded), if knowledge of the bomb would cause viewers to look back to the car more quickly. To perform this analysis, 8 time points were identified during which the car was completely occluded from view. This could be due to the car going off screen, driving behind occluding objects on the screen, or having objects on the screen moving in front of the car. For these 8 instances, starting from the exact time point at which the car was again visible (disocclusion), a 3 second time window was created. This time window was used to test the amount of time it took for participants to fixate the car after each disocclusion for each group. If a participant did not refixate the car within the 3 second time window they were not included in the analysis for that occlusion. For each disocclusion time point a t-test was performed. Figure 5 shows the average latency to refixate the car after each disocclusion. None of the eight disocclusions showed significant differences in the latency to refixate the car. Disocclusions 1-6 and 8 had p's > .05. Disocclusion 7 only had three participants refixate the car in the 3 second time window, all of whom were in the No-context group, which meant the t-test could not be run. Also, as seen in the figure, the group that fixated the car more quickly differed between disocclusions, and for four of the disocclusions fewer than

half of the participants fixated the car within the 3 second time window, producing unbalanced samples for each group.

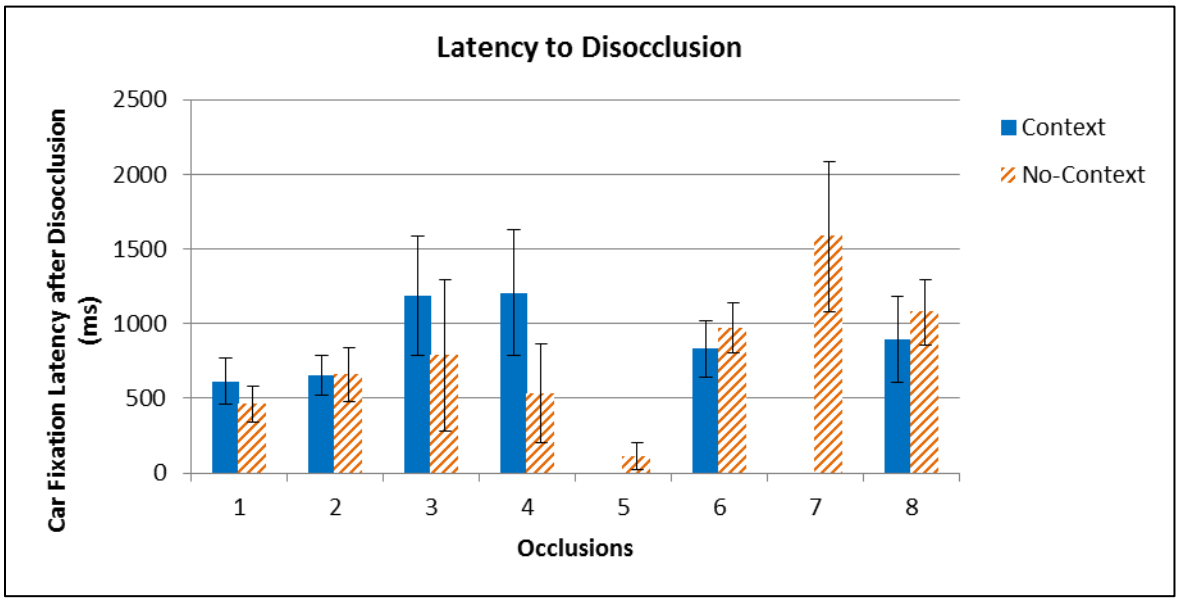


Figure 5. Time taken (with standard error bars) for participants to fixate the car after it became disoccluded (i.e., came back into view after having been out of view) as a function of context condition.

The lack of measured differences in eye-movements due to the comprehension manipulation is similar to the results of the *James Bond Moonraker* study (Loschky et al., 2015). One simple reason for the similar results between Loschky et al. (2015) and this experiment could be that the *Touch of Evil* film clip did not reduce the overall attentional synchrony compared to the *Moonraker* clip. Direct comparison between the participants in the two studies on measures such as gaze similarity is not possible. This is because the clips have differences in features that could have an effect on the measures. These differences include but are not limited to the aspect ratio of the clips, the resolution of the clips, and overall compositional differences. However, it is possible to descriptively compare the overall amount of gaze clustering between the 2 film clips. Two statistics of gaze clustering were used for comparison; the sum weighted covariance and the number of clusters. Both of these are statistics available from CARPE (Mital et al., 2010), the program used to process the data for the gaze similarity analyses. The sum weighted covariance of gaze is calculated for each frame of each film for each participant condition. The covariance measure is the sum of the covariances of the optimal number of

clusters used to describe the distribution of the gaze during each frame. The "weighted" component of the measures indicates that the covariance measure gives more weight to clusters that are composed of more gaze points (see Mital et al. (2010) for more details). This is essentially the amount of spread in gaze for each frame, while controlling for the number of clusters of gaze and how many participants are in each cluster. The number of clusters statistic is the minimum number of regions in the frame that can best describe the distribution of all participants' gaze. Higher values for both measures indicate greater gaze dispersion, and thus lower attentional synchrony. Figure 6 shows that both of these measures of gaze dispersion were higher for *Touch of Evil*, indicating that it produced lower attentional synchrony than the *Moonraker* clip. Therefore, the motivation for choosing the *Touch of Evil* film clip was warranted, as it should have produced less tyranny of film than the *Moonraker* clip. Nevertheless, this stands in contrast to the lack of effect of strong comprehension differences on viewers' eye-movements.

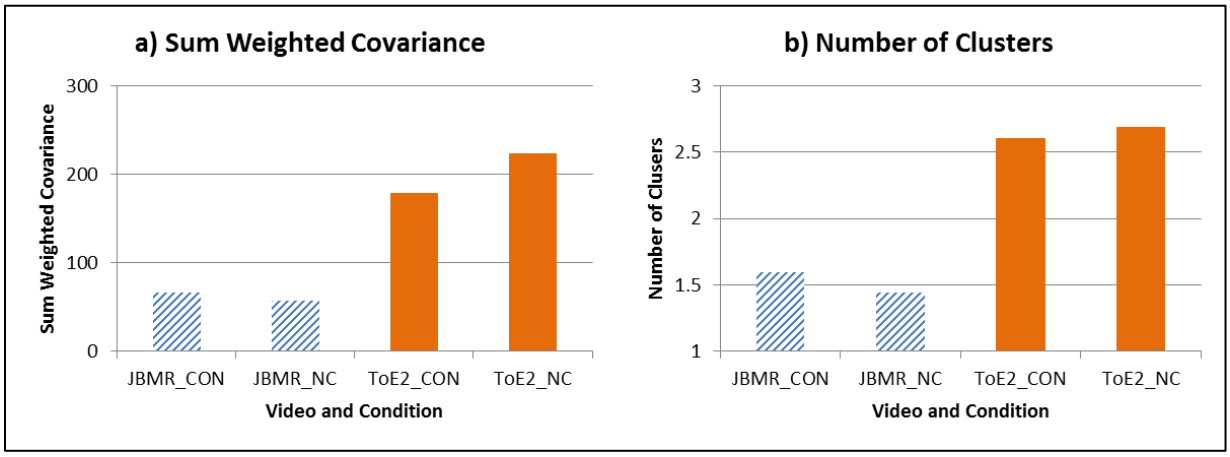


Figure 6. a) Sum weighted covariance of gaze averaged over all frames. Values indicate the amount of space occupied by gaze, with smaller values signifying tighter clustering of gaze, and higher values signifying greater gaze dispersion. b) The average number of gaze clusters across all frames.

Discussion

In Experiment 2, the strong effect of the Context manipulation on comprehension from Experiment 1 was replicated. However, Context had only one small effect on eye-movements. Saccade lengths were longer for participants in the Context condition that made the inference about the bomb. All other eye-movement measures showed no effect, including attentional

synchrony, the frequency of looking at the car with the bomb in it, or the latency to refixate the car after it came back into view (after being occluded). These results therefore mostly support the tyranny of film hypothesis. This is despite the fact that, as shown in Figure 6, the *Touch of Evil* film clip produced less gaze clustering than the *Moonraker* clip, as we had predicted due to the *Touch of Evil* clip lacking cuts and including numerous objects in the film frame to look at. Thus, despite having overall lower gaze clustering and what would appear to be a stronger manipulation of viewers' comprehension than in the previous *Moonraker* study, viewers' eye-movements still did not reliably differ based on their understanding of the clip. The greater amount of dispersion in *Touch of Evil* shown through the low gaze clustering does mean it is harder to find group differences in gaze similarity, because participants are exploring more of the screen. However, the region of interest analysis should be more sensitive to potential location based eye-movement differences.

There are some potential problems with both our manipulation of comprehension and our measure of it in our first set of experiments. First, as discussed earlier, in both Experiments 1 and 2, a single participant in the No-context condition who heard audio was able to make a bomb-relevant inference without having seen the bomb at the beginning of the clip, or having seen the movie before. We hypothesized that this was due the fact that, towards the end of the clip, a character in the car mentioned hearing a ticking noise, thus enabling those participants to guess that there was a bomb in the car, and so allowing them to generate a bomb-relevant inference a short time later when asked "what will happen next?" Furthermore, we hypothesized that this would explain why twice as many viewers in the Context + Audio condition generated the inference compared to the Context + No-audio condition in Experiment 1. Namely, the difference could be due to half of the viewers who made the inference in the Context + Audio condition having forgotten about the bomb and being reminded when they heard the ticking comment shortly before the clip ended. This was the motivation for measuring working memory in the next experiment. Since participants higher in working memory tend to have better narrative comprehension, it is predicted that participants with higher working memory will be more likely to make the bomb inference (i.e., less likely to forget about the bomb). If roughly half of the Context condition participants end up forgetting about the bomb during the clip, but are then being reminded by the "ticking" comment near the end of the clip, then those viewers would not have maintained the bomb in their mental model throughout the three minute clip, and

so their eye-movements should be more similar to participants in the No-context condition. Therefore, if roughly half of the participants in the Context condition were likely to have forgotten about the bomb, this could potentially explain the lack of effect of Context condition on eye-movements in Experiment 2. To remove the statistical noise in the inference data potentially introduced by the "ticking" comment, all audio was removed from the second set of experiments. Thus, participants that make the bomb related inference at the end of the clip will need to have maintained it in their mental model throughout the duration of the clip.

An additional potential problem with the first set of experiments was that the No-context condition and the Context condition both show the car at the beginning of the scene, and in particular a couple getting into the car. First mentioned entities have a special status in mental models for narratives (e.g., Gernsbacher, 1990). As such, the car was likely prominent in the mental models for both the Context and No-context conditions, which may have led to similar eye-movements in both conditions, regardless of whether they had knowledge of the bomb.

Finally, a third and more general problem is that our measure of comprehension, namely whether or not a viewer generated a bomb-relevant predictive inference at the end of the clip, was unidimensional. It is possible that such a measure, taken after the viewers' have finished watching the film clip and their eye-movements recorded, overestimates the differences in viewers' comprehension, or perhaps the predictive inference measure is simply insensitive to comprehension processes that affect eye-movements. This suggests that other richer measures of comprehension may be needed to find comprehension effects on film viewers' eye-movements.

Chapter 4 - Experiment 3: Event Segmentation and Working Memory with New No-Context Condition

The second set of experiments (Experiments 3 & 4) were designed to deal with all of the potential reasons the first set of experiments found minimal differences in eye-movements despite the context manipulation. The experimental changes were both in terms of the context manipulations and the comprehension measures.

Context manipulations

To address the potential problem that viewers in the Context condition may forget about the bomb until they hear the "ticking comment" near the end of the clip, but then subsequently report a bomb-relevant predictive inference, the clip was shown without audio. With this manipulation only viewers in the Context condition who maintain the bomb and its potential destructive causal implications in their mental model throughout the clip should draw a bomb-relevant inference at the end. Conversely, those viewers who forget about the bomb while watching the clip should not be reminded by the "ticking comment" near the end of the clip, since it is only in the audio, and should therefore not draw a bomb-relevant predictive inference. This should produce a stronger difference in the mental models of those participants who draw the inference in the Context condition, and those who did not, particularly in the No-context condition.

Furthermore, in order to address the potential issue that viewers in the No-context condition may have treated the characters in the car as protagonists, and therefore paid close attention to the car even though they did not know it contained a bomb, a new No-context condition was used. In the new No-context condition, viewers began watching the clip only after the walking couple entered the street and the car was off-screen (Figure 1; Image 5 marked "No-context: Exp. 3 & 4"). Thus, viewers in the new No-context condition should treat the walking couple as the protagonists. Viewers in the Context condition started watching the clip from the beginning as before.

Comprehension measures

In the first set of experiments the only comprehension measure used was predictive inference generation, which focuses on high-level comprehension relatively globally at the end of the clip. On the other hand, eye-movements are an on-line measure of attention. Experiment 3 therefore used event segmentation as an on-line comprehension measure throughout the entire film clip (Newtson, 1973; Zacks, Speer, & Reynolds, 2009). Event segmentation involves having participants press a button any time they perceive a new event occurring. Importantly, however, eye-movement and event segmentation measures were not run simultaneously, as the manual task of pressing a button for event segmentation will likely have an effect on eye-movements (Eisenberg & Zacks, 2016). However, it has been demonstrated that event segmentation is a naturally occurring process that aids comprehension (Hard, Tversky, & Lang, 2006; Zacks & Tversky, 2001), and that participants commonly show high inter-subject and test-retest reliability (Newtson, 1973; Speer, Swallow, & Zacks, 2003; Zacks, Speer, Swallow, & Maley, 2010).

Additionally, in the previous experiments it was shown that roughly half of the participants in the Context condition who see the bomb do not end up making a bomb-relevant predictive inference at the end of the film clip. In an attempt to explain this, participants in Experiment 3 completed a series of working memory measures: operation span (OPSAN, Turner & Engle, 1989) reading span (RSPAN, Daneman & Carpenter, 1980), and counting span (CSPAN, Case, Kurland, & Goldberg, 1982). Reading and comprehension research has shown that working memory span is an individual difference that predicts comprehension abilities such as making predictive inferences (Allbritton, 2004; Calvo, 2001, 2005; Daneman & Carpenter, 1980; Daneman & Merikle, 1996; Just & Carpenter, 1992; Linderholm, 2002; Rai et al., 2014; Rai et al., 2011; St George et al., 1997). Although there are different processes that contribute to comprehension in reading and visual narratives, there should also be overlapping comprehension processes (Loughlin & Alexander, 2012; Magliano et al., 2013). Based on this, we predicted that participants in the Context condition with higher working memory scores would be more likely to make a bomb-relevant inference than those in the Context condition with lower working memory scores.

Method

Participants

A total of 81 students enrolled in an introductory psychology course at Kansas State University participated in Experiment 3 for course research credit, and were pseudo-randomly assigned to either the Context condition (n = 41) or the No-context condition (n = 40). Participants took the FrACT visual acuity test (Bach, 2006), and all had 20/30 or better corrected or uncorrected vision. Only participants who had previously participated a working memory study with the OSPAN, RSPAN, and CSPAN were given the possibility of signing up to participate in this experiment. Despite this, working memory scores were only available for 31 participants in the Context condition, due to participants either not completing the working memory study or issues of matching participants between the two studies.

Stimuli

The same opening scene from *Touch of Evil* was used in Experiment 3, but no audio was presented with the clip. The clip used for the Context condition was identical to that used in Experiments 1 and 2. The new No-context condition saw a different version of the clip that started 1 minute and 49 seconds into the opening scene, at a point when the walking couple was shown alone on the screen, with the car off-screen.

Participants viewed the clip in groups of up to four. Each participant viewed the clip on an individual 17 inch Samsung SyncMaster 957 MBS monitor, with a chin rest used to maintain a constant viewing distance of 53.34 cm, a visual angle of 27.06° x 18.18°, and an image resolution of 1080 x 720 pixels. The monitors were set to refresh at 60Hz, to ensure there were no dropped frames with the video playing at 30 frames per second.

Procedure

To learn the process of event segmentation participants were first given practice with the task using an example 83 second video of a person folding laundry. They were instructed to press a button whenever they perceived "new" events that were "natural" and "meaningful." After the practice video the instructions were repeated, and participants were able to begin the experimental video clip by pressing a button. As in Experiments 1 and 2, once the video was

complete, participants were asked, "What will happen next?" This was followed by a question of whether they had seen the film before, and, if so, what its name was. Two participants were removed from the analysis because they had seen the clip before.

Results

Predictive inference

The inference coding in this experiment followed the same procedure as the previous two. Two coders independently identified the predictive inference participants made as either relating to the bomb, or not. The inter-rater reliability was high (*Kappa* = .962, $p < .001$). Remaining discrepancies between the two coders were resolved through discussion (there was only one inference in this experiment that needed discussion).

A chi-square analysis showed that, as before, participants in the Context-condition were far more likely to make the predictive inference about the bomb than participants in the No-context condition (X^2 (1, $N = 82$) = 18.932, $p < .001$). Thus, the *jumped-in-the-middle* Context manipulation again strongly affected participant's mental model of the film clip, as measured by the predictive inference task.

Working memory and inference generation. To test whether participants with better working memory scores were more likely to make a bomb-relevant predictive inference, a *t*-test compared working memory scores for those participants in the Context condition who did or did not make the inference. Of the 41 participants in the Context condition, only 31 had completed all components of the working memory experiment and could be matched up to their results in the event segmentation experiment (12 participants in the group of 31 made the inference). As shown in Figure 7 the composite working memory score across all three measures showed a main effect of working memory on inference generation (t (29) = 2.310, $p = .028$, $d = .918$). Participants that made an inference about the bomb recalled proportionately more working memory items (.84, SD = .08) than those participants that did not make the inference (.71, SD = .19). For the individual working memory measures, the CSPAN had the largest effect size (t (29) = -2.320, $p = .028$, $d = .983$) and highest overall performance for both groups (Inference Mean = .92, SD = .05; No-Inference Mean = .78, SD = .19), the OSPAN was in the middle (t

(29) = 2.010, p = .054, d = .776) with Inference group performance at .79 (SD = .15) and the No-Inference group at .63 (SD = .23),and the RSPAN had the lowest (t (29) = 1.763, p = .089, d = .703) with inference group performance at .83 (SD = .08) and No-Inference group performance at .71 (SD = .21). Therefore, it seems that despite differences among the three working memory measures, the relationship between working memory and inference generation from the film clip is stable. Thus, our predictive inference measure of high-level film comprehension shows the same sort of relationship to working memory as has been shown in previous studies of inference generation in reading comprehension (Allbritton, 2004; Calvo, 2001, 2005; Daneman & Carpenter, 1980; Daneman & Merikle, 1996; Just & Carpenter, 1992; Linderholm, 2002; Rai et al., 2014; Rai et al., 2011; St George et al., 1997), lending support to the idea that the predictive inference measure is indeed measuring film comprehension processes.

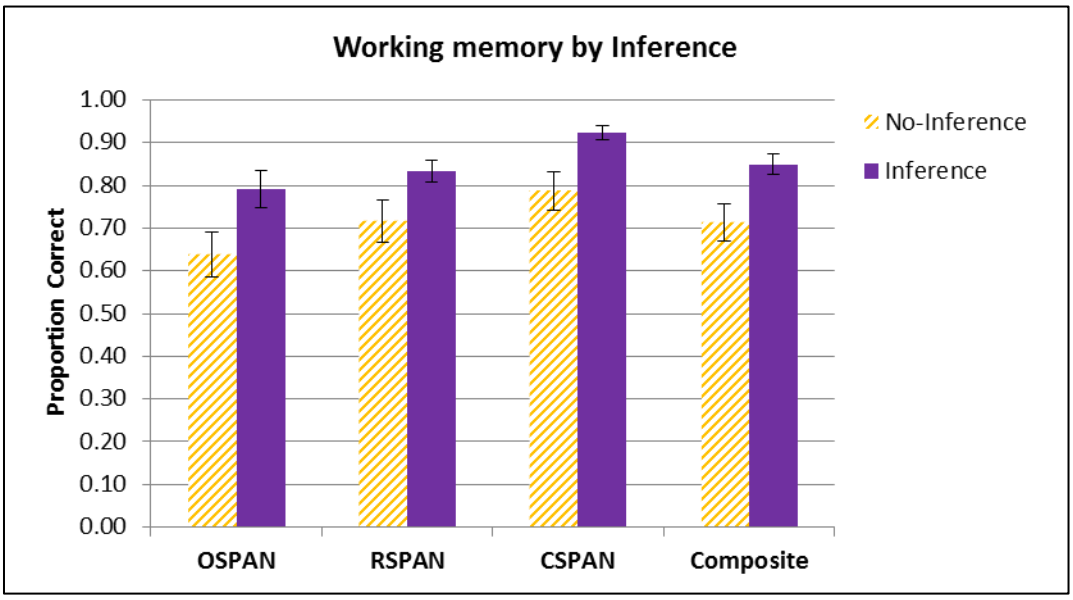


Figure 7. Proportion of working memory items recalled accurately (with standard error bars) for all working memory measures by presence/absence of bomb inference.

Event segmentation

Proportion of events. To test for a general difference in event segmentation throughout the portion of the clip that both conditions saw, the proportion of participants who identified events within 3 second bins was calculated. This first analysis tested whether there was a

difference in the proportion of events identified by the Context conditions from the first appearance of the walking couple to the point they kiss at the very end of the clip. For this analysis the clip was first divided into three second bins. Next, for each participant, it was calculated whether they pressed their button to identify an event for each of the bins. This gave an overall distribution of when events were identified by each participant throughout the film clip, and then these were aggregated to give an overall group distribution for events in the clip. No-context participants indicated a significantly higher proportion of events (No-context = .32) than those in the Context condition (Context = .20)(X^2 (1, N = 44) = 44.00, p = .021; Eta = .350) (Figure 8). This provides converging evidence, here measured in an on-line task, that the context manipulation creates differences in film comprehension, with the Context participants generating fewer boundaries than the No-context group. The fact that the Context group identified fewer boundaries may appear counterintuitive since they have a better understanding for the narrative. One explanation for this is that better understanding should allow for better prediction of upcoming events, and event segmentation typically occurs when there is an error in the prediction of an event (Reynolds, Zacks, & Braver, 2007). Alternatively, the No-context group may identify more events because they are tracking more events occurring outside of the primary narrative about the bomb (e.g., the introduction of new characters and events that are irrelevant to the bomb narrative).

Additionally, since those with higher working memory showed a greater probability of making the inference for those in the Context group, the role of working memory on segmentation rate was also tested using regression with the same Context group participants. Drawing from the previous result, it was expected that those higher in working memory who had better comprehension would identify fewer events overall. This is because having higher working memory capacity may require less mental model updating. The working memory composite was entered as a predictor of the proportion of events overall. No relationship was found between working memory and the proportion of events identified (B = -8.07, R^2 = .068, F (1, 27) = 1.977, p = .171), although the relationship trended in the predicted direction of participants with higher working memory identifying fewer events.

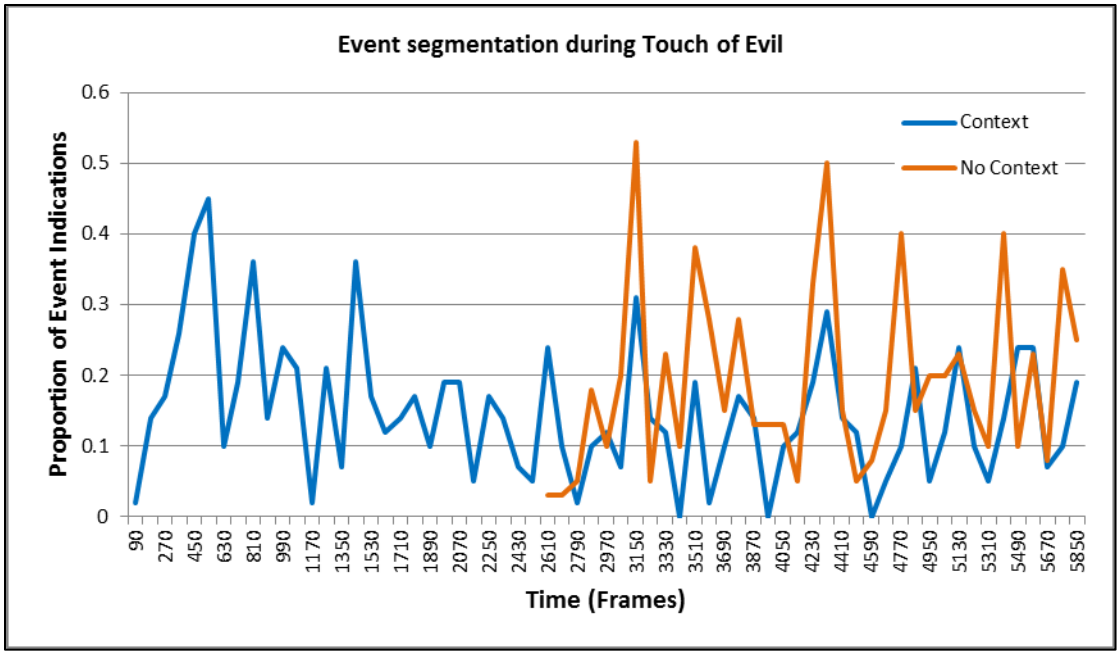


Figure 8. Proportion of new events indicated within 3 second bins by context condition.

Proportion of events: First appearance of car. The new No-context condition was intended to create differences in the perceived protagonists based on which entities were seen first. Thus, for the new No-context condition, the first appearance of the car should be a critical viewing period in which the car is added the mental model. In the Context condition, participants already know about the car and where it is, and should be more likely to predict it will reappear where it does. To test this hypothesis, we created an 8 second time period, which began when the car first appeared on the screen for the new No-context condition, and ended when the car was occluded (as described in Experiment 2, *Latency to Fixate the Car after Disocclusion*). We then calculated the proportion of events perceived during that 8 second period in each context condition, and found similar results to those for the entire shared viewing period. Participants in the No-context condition indicated a significantly higher proportion of events (No-context = .0536) than those in the Context condition (Context = .0397) (X^2 (1, N = 13) = 13.00, p = .043, Eta = .303), though the overall rate of event perception during the 8 second period was relatively low for both groups.

Segmentation agreement. An alternative method of scoring the event segmentation data is in terms of the proportion of segmentation differences, which allows us to ask whether there

are differences in where participants segment throughout the clip. For this, we used Zacks' (1999) segmentation agreement scoring method. This analysis marks events within one second bins and creates a segmentation baseline that each participant is compared to create an agreement score for each participant in the form of a scaled correlation.

The first agreement analysis used all participants to create an overall event structure for the shared viewing portion of the clip. In this analysis, we compared each condition against all other viewers, to determine whether there were general similarities in event segmentation across groups. When comparing the scaled correlations for each participant, we did not find significant differences in agreement scores between the Context (.515) and No-context (.526) conditions, (t (76) = -.308, p = .759; d = .071). This indicates that at a gross level, both groups had a similar level of agreement, but it does not take into account the potentially different event structures for both groups.

Our second agreement analysis mirrored the gaze similarity eye-movement analysis by comparing both groups to a baseline set by the Context condition. In other words, the first agreement analysis used all participants to create the comparison group regardless of condition. The second analysis only used the comprehension group to create the comparison group. This allowed for a test of whether the No-context group's event structure was significantly different from the Context group. For this, we first calculated the event structure identified by the Context condition, and then individually compared the Context and No-context condition participants to the Context condition. Using the Context group as the comparison group, a t-test of the scaled agreement scores showed a main effect of Context on event segmentation agreement (t (76) = 2.898, p = .005; d = .664). This indicates that there were differences in where Context and No-context participants identified events within the shared viewing period of the clip. This, again, provides converging evidence of on-line differences in comprehension as a function of context condition[5].

Lastly, the relationship of working memory and segmentation agreement was tested based on the known relationship of working memory and comprehension (Daneman & Merikle, 1996). High segmentation agreement typically indicates better comprehension (Zacks et al.,

[5] The agreement analysis could not be run for the first appearance of the car for the No-context group. The 8 second viewing period did not have enough events identified in either condition to allow for a reliable event structure to be created for statistical comparison.

2009; Zacks & Tversky, 2001), thus it is expected that those high in working memory should also show greater agreement. The working memory composite was entered as a predictor of segmentation agreement. No relationship was found between working memory and segmentation agreement (B =.01, R^2 = .0002, F (1, 27) = .006, p = .937). It is somewhat surprising that there is not an effect of working memory on segmentation agreement, because it is thought that segmentation is directly related to the updating of information in working memory (Kurby & Zacks, 2008). Although, the lack of the difference could be similar to the reasoning for the null effect of eye-movements, which is that although certain participants forget about the bomb they still follow the actions of the same characters and events.

Discussion

Experiment 3 had two primary goals. First, it was designed to address the concern from Experiments 1 and 2 that the predictive inference might not have adequately measured comprehension at the end of the clip. To the extent that we have not been adequately measuring comprehension, it could explain the lack of effects of comprehension on eye-movements. For this reason, we used the event segmentation task as a qualitatively different on-line measure of film comprehension. The event segmentation results showed that the context manipulation resulted in participants identifying events at different points throughout the film clip, and the No-context group was more likely to identify events throughout. These results indicate that the new No-context manipulation had the type of effects on comprehension that we would expect, providing converging evidence for the validity of our comprehension manipulation.

The second goal was to help explain the fact that roughly half of the participants in the Context condition failed to generate a bomb-relevant predictive inference. It seems surprising that so many viewers in the Context condition would fail to draw bomb-relevant inferences at the end of the clip, despite the bomb's central importance to the narrative in the opening shot of the film. We therefore tested the hypothesis that viewers in the Context condition with lower working memory would be less likely to generate a bomb-relevant inference at the end of the clip, and found evidence strongly consistent with it. We suspect that this result was due to the fact that after first seeing the bomb at the beginning of the clip, there were no further cues to the existence of the bomb throughout the remainder of the opening shot, particularly after the exclusion of the audio track, which included the "ticking comment" near the end of the clip.

Thus, viewers with lower working memory would have a harder time maintaining the bomb in their mental model of the narrative in the face of other competing salient narrative events (e.g., the introduction of other locations, characters, and their actions). By showing that working memory limitations were strongly related to the process of drawing the target predictive inference at the end of the clip in the Context condition, it connects our results to established theoretical explanations of higher-order processes in reading comprehension, and thus lends credence to our claims to be validly measuring comprehension of the film clip through use of the predictive inference task. It also provides a richer description of viewers' comprehension processes involved in understanding the particular film clip used in our study. Finally, to our knowledge, this study is the first to have extended the working memory span to inference generation findings from the realm of reading comprehension to visual narrative comprehension (but see Magliano, Larson, Higgs, & Loschky, 2015).

Finally, the role of working memory in event segmentation frequency and agreement was tested. The two main reasons for this were that segmentation of events is thought to occur at the updating of event model information in working memory (Zacks et al., 2009; Zacks & Tversky, 2001), and working memory in general has been shown have a relationship with comprehension processes (Allbritton, 2004; Calvo, 2001, 2005; Daneman & Carpenter, 1980; Daneman & Merikle, 1996; Just & Carpenter, 1992; Linderholm, 2002; Rai et al., 2014; Rai et al., 2011; St George et al., 1997). Despite this, no effect of working memory on event segmentation was found.

The large number of effects of the manipulation of comprehension motivates a new eye-tracking experiment in a number of ways. Foremost, the effects show that the new No-context condition and the removal of audio are creating clear distinctions between the mental models for each condition as measured by inference generation, and this measure has less noise due to the removal of the "ticking" comment. Additionally, the event segmentation is showing online effects of the manipulation, which may indicate that there should also be an effect on the online measure of eye-movements.

Chapter 5 - Experiments 4a and 4b: Eye-Tracking with New No-context Condition and Map Task

Having created the new No-context condition, and establishing that it created strong effects on comprehension of the film clip in Experiment 3, we carried out Experiments 4a and 4b. Experiment 4a was designed to determine what effects, if any, the new No-context condition would have on eye-movements. Experiment 4b introduced a new condition to test for the effect of task on comprehension and eye-movements.

Experiment 4a

Method

Participants

Data was collected from 201 students enrolled in an introductory psychology course at Kansas State University for course research credit. Data from 8 participants were dropped because of program errors during data collection, for not completing the questions at the end of the experiment, or for having participated in an earlier experiment using *Touch of Evil*. Data from the remaining 193 participants (Age: M = 19.5, female = 59.2%) were included in the analyses. Participants were pseudo-randomly assigned to either the Context condition (n = 131) or the new No-context condition (n = 62 participants) with the constraint that we have roughly twice as many participants in the Context condition, based on the assumption (from Experiments 1-3) that roughly 50% of them would fail to generate a bomb-relevant predictive inference at the end of the film clip.

Stimuli

The video stimuli used in this experiment were the same as those used in Experiment 3, namely all video clips were presented with No-Audio. As in Experiment 2, stimuli were presented on a 17" ViewSonic Graphics Series CRT monitor (Model G90fb), and using a chin and forehead rest, there was a fixed viewing distance of 60.96 cm. The screen subtended 21.42° x 16.10° of visual angle.

Procedure

Experiment 4a's procedures were identical to those of Experiment 2 with the exception of the above-noted changes to the No-context condition, and the lack of audio in the film clip.

Data analysis

The procedure for predictive inference coding was identical to Experiment 1, and was carried out by the same two research assistants. Once the research assistants had coded each participant response, inter-rater reliability was tested, and shown to be high using Cohen's *Kappa* = .954, $p < .001$. Any remaining discrepancies between the two coders were resolved through discussion such that each response was coded either as '0' or '1.'

Results

Predictive inference

Inference coding was used for two purposes in this experiment. First, for participants in the Context condition, it allowed us to do analyses based on comprehension. Second, it allowed us to exclude any participants in the No-context condition that made the predictive inference about the bomb. The procedure here was the same as in the previous 3 experiments. Overall, we replicated the results of the previous experiments with the Context condition more likely to make a bomb-relevant inference (X^2 (1, N = 193) = 46.39, $p < .001$) (Eta = .490). Almost exactly half of the participants in the Context condition made a bomb-relevant inference (65 participants made the inference and 66 did not), consistent with Experiments 1-3. No participants in the No-context condition made the inference. Thus, data in all the following analyses are in terms of three groups: the Context participants that made the inference (Context + Inference), the Context participants that *did not* make the inference (Context + No-inference), and the No-context group (none of whom made the inference).

Eye-movements

Fixation durations and saccade lengths

Data cleaning followed the same procedure as Experiment 2. The effects were reversed when compared to Experiment 2. There was an effect of group on fixation duration (F (2, 191) = 3.79, p = .024, η_{p2} = .038). Tukey HSD post hoc comparisons indicated that the Context + Inference group had the shortest average duration of 367 ms (SD = 57 ms), which was significantly shorter (p = .026) than the No-context group at 398 ms (SD = 82 ms). The Context + No-inference group was not different from either of the other groups (378 ms; SD = 70 ms). There were no differences in average saccade length (F (2, 191) = .905, p = .406). The Context + Inference group averaged 4.89° of visual angle (SD = .61°), for the Context + No-inference group it was 4.65° (SD = .74°), and for the No-context group it was 4.69° of visual angle (SD = .82°).

Shorter fixation durations for the Context group that made the inference match the relationship between fixation durations and comprehension during reading. Specifically, fixation durations tend to be shorter when a person has a better understanding for what they are reading (Rayner, 1998). In the current experiment participants in the Context condition that made the inference show a high level of comprehension for the narrative and shorter fixation durations. Further exploration is needed to identify why in Experiment 2 there was a medium sized effect of context on saccade lengths, and then a medium size effect on fixation durations in Experiment 4. Either the use of the new No-context or the removal of the audio track could have made the saccade length effect smaller for Experiment 4, and created the fixation duration effect.

Attentional synchrony

To perform the analyses for Experiment 3 the same data pre-processing that was employed in Experiment 2 was used. CARPE (Mital et al., 2010) was used to down sample the raw eye-movement data. After processing this data, a reference probability distribution was calculated for the Context with inference group. The eye-movement data for each participant in each group were then compared to the reference distribution to calculate the gaze similarity between groups.

As with Experiment 2, the first run was an omnibus test for gaze differences across the viewing of the critical portion of the film clip (the same 1 minute and 49 seconds of the clip that

both groups saw). During this critical portion of the clip, there were no group differences (F (2, 190) = 0.05, p = .955). As in Experiment 2, Figure 9 shows that the lines indicating gaze similarity for each group are nearly identical, indicating that the tendency for viewers to look at the same places at the same times was the same across groups. Again, a shuffled baseline was included for comparison for the shared viewing period. Overall, the gaze similarity of all three experimental groups' was generally above the shuffled baseline, indicating that the film was guiding eye-movements, creating the tyranny of film. Including the shuffled baseline in the ANOVA was statistically significant (F (3, 254) = 55.23, p < .001, η_{p2} = .395), with all experimental groups higher than the baseline.

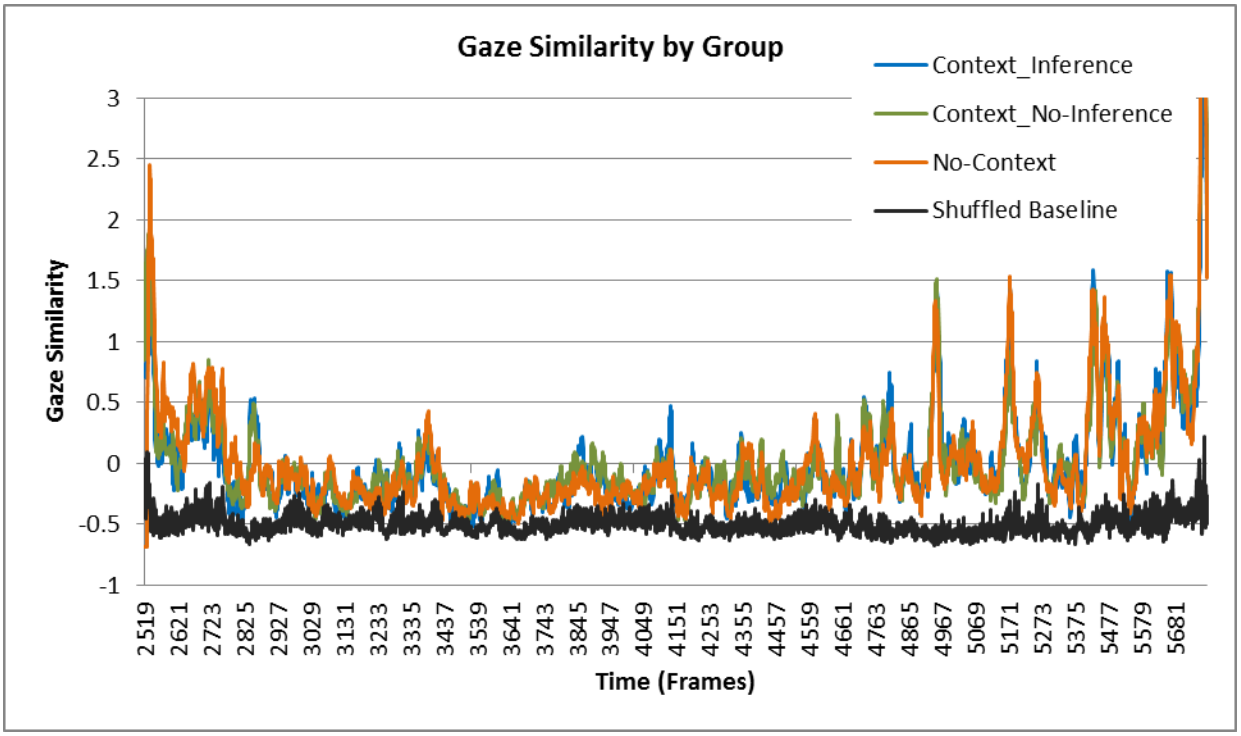


Figure 9. Similarity of gaze by context condition across the shared viewing period of the clip that starts on frame 2519. Gaze similarity is expressed as a z-score probability relative to the context condition and inference made (Context + Inference [Blue], Context + No-Inference [Green], No-context [Orange], and Shuffled Baseline [Black]). Larger values indicate greater attentional synchrony. The shuffled baseline indicates chance level gaze similarity for the clip.

What these results indicate is that viewers' understanding of the clip did not influence their attentional synchrony. Even when viewers knew about the bomb or thought the car and the

couple in it were the main characters of the film clip, they viewed the clip similarly to viewers who did not. However, we cannot explain this lack of differences between groups in terms of the film clip failing to guide eye-movements since overall gaze similarity was clearly well above chance for all three groups.

Region of interest

The same region of interest data pre-processing as in Experiment 2 was used. The car was identified as the region of interest, and we tested whether our three viewing groups differentially looked at it. The region of interest analyses again started using the entire shared viewing period, and then a predetermined time point of interest based on the manipulation of protagonist.

The omnibus region of interest analysis started the first time the car appeared on the screen in the No-context condition, 1 minute and 57 seconds into the film clip. Overall, there were no significant differences between the three groups in how often they fixated the car during this viewing period (F (2, 190) = 1.07, p = .345) (Figure 10). This is in line with the gaze similarity analysis, indicating that the manipulation of both knowledge of the bomb and the protagonists had no effect on viewers' overall likelihood of looking at the car.

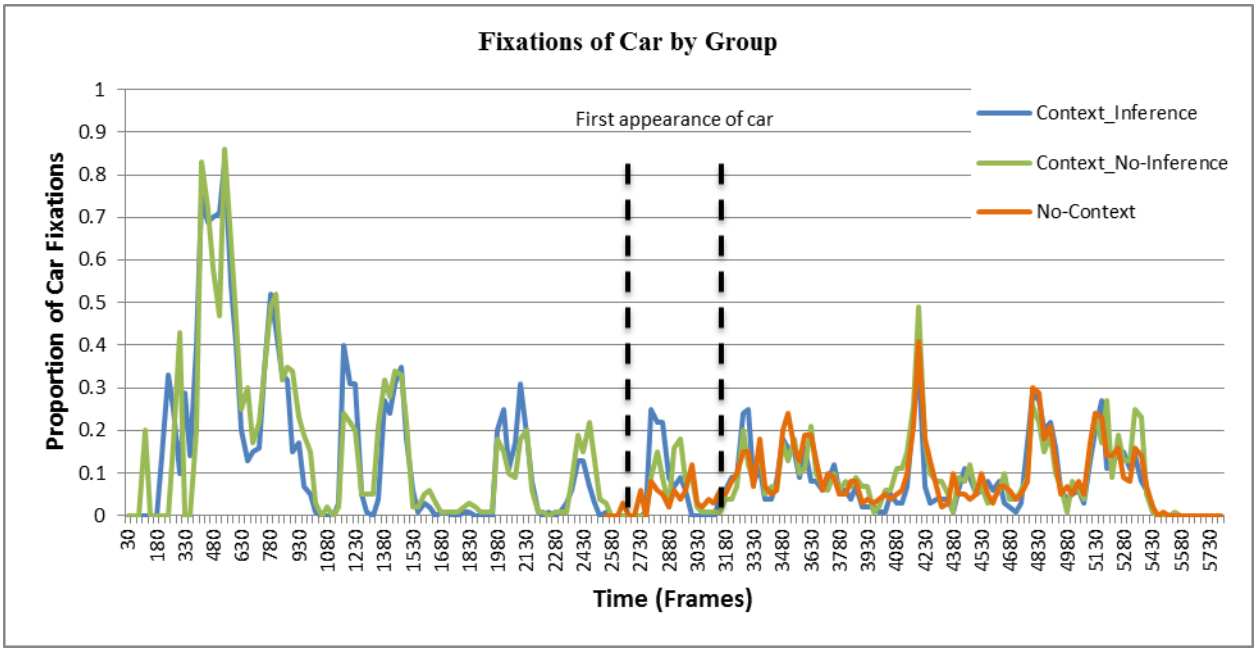


Figure 10. Proportion of participants fixating the car by context condition and inference throughout the film clip (Context + Inference [Blue], Context + No-Inference [Green], and No-context + No-Inference [Orange]). The higher the value the more participants looking at the car. First appearance of car for No-context group marked with dotted lines.

We next carried out a more specific region of interest analysis to probe a critical time period when the manipulation of protagonist might be expected to have an effect on fixations of the car. This time period was the same 8 second period used in the event segmentation analysis when the car with the bomb was first seen by participants in the No-context condition. At that point in the narrative, the walking couple walks past the car, which is not moving because a crowd of pedestrians has blocked the street. For viewers in the No-context condition, the car should have no particular importance, but for viewers in the Context condition, it should already be an integral component of their mental model, regardless of whether the bomb is still active in their mental model or not. Thus, in line with the "eye-mind hypothesis" (Just & Carpenter, 1980; Reichle et al., 1998; Reilly & Radach, 2006), the mental model hypothesis would predict that the Context condition viewers would be more likely to look at the car than the No-context condition, at least initially. To test this prediction, we created an 8 second time window from when the car first appeared (frame 2758) and right before it was briefly occluded from view again (frame 3028), and we then measured the proportion of fixations of the car for each group. A one-way (Group: Context + Inference vs. Context + No-inference vs. No-context) between

subjects ANOVA found a main effect of viewing group on proportion of fixations on the car during the pre-specified 8 second time window (F (2, 190) = 3.93, p = .021, η_{p2} = .04). As illustrated in Figure 10, the Tukey HSD procedure indicated that viewers in the No-context group were significantly less likely to fixate the car (M = .054, SD = .078), than those in the Context + Inference group (M = .098, SD = .086) (p = .02). There was also a non-significant trend (p = .06) for viewers in the Context + No-inference group to fixate the car less than viewers in the Context + Inference group (M = .093, SD = .122). This provided the first support for the influence of the mental model on predictable gaze behavior in the *Touch of Evil* film clip. Viewers who knew about the bomb and already had the car in their mental model were more likely to fixate the car during its first appearance within the critical period than participants who had not indexed the car (and the couple in it) as important agents in the narrative. We will refer to this as evidence of *the agent effect*, as it seems to be due to whether viewers treat an entity in the narrative as an "agent" or not.

In general, experiment 4a showed support for the tyranny of film with there being no overall differences in gaze distribution across the difference inference conditions. The fixation duration effect may indicate that Context + inference participants are under less cognitive load than No-context participants, which is consistent with what is expected for participants with better comprehension (Rayner, 1998). The agency effect, reported for the ROI analysis demonstrates the potential for gaze to be influenced by an object's relevance to the viewer's mental model, but the fleeting nature of this effect suggests that the motivation for such top-down control may need to be stronger and more deliberate during film viewing than has been demonstrated in static scene viewing (DeAngelus & Pelz, 2009; Smith & Mital, 2013; Yarbus, 1967) due to the bottom-up tyranny of film. With this overall lack of an effect, Experiment 4b was conducted to test whether it was possible to get a strong effect with our stimuli and measures. This tests whether the previous experiments are showing a true lack of an overall effect of comprehension on eye-movements, or if there may be a problem with our stimuli or measures.

Experiment 4b: Map Task

In Experiments 2 and 4a, we found that participants who view the clip with different understandings show surprisingly few eye-movement differences, which generally supports the

tyranny of film hypothesis. This suggests that differences in narrative comprehension during film viewing have little effect of visual attention. This may be due to differences in comprehension not creating differences in the objects of interest in the narrative similar to Taya et al. (2012). However, previous work has shown higher level cognition has large, online effects on eye-movements during scene viewing. Classic work on eye-movements in static scenes has shown cognitive effects of viewing task on eye-movements, such as looking at a picture and trying to determine what the people's ages are, or how wealthy they are (DeAngelus & Pelz, 2009; Yarbus, 1967) and more recent studies comparing, for example, scene memory versus object search tasks (Foulsham & Underwood, 2007; Henderson, Brockmole, Castelhano, & Mack, 2007; Henderson, Shinkareva, Wang, Luke, & Olejarczyk, 2013). These effects have also been shown more recently in natural film (i.e., unedited real-world video) such as trying to determine the location depicted in a video (Smith & Mital, 2013) and edited narrative film (Lahnakoski et al., 2014). The latter study showed viewers a clip from "Desperate Housewives" twice, and for each viewing manipulated participants' cognitive perspective, either that of a detective or of an interior decorator. We consider adopting such a cognitive perspective to be a viewing task manipulation.[6] The Lahnakoski et al. (2014) results showed that when adopting the detective perspective, viewers looked at the main characters more (in the center of the screen, making shorter saccades), but when adopting the interior decorator perspective, they looked at the background setting more (on the edges of the screen, making longer saccades). Interestingly, after the experiment, some participants mentioned that when they adopted the interior decorator perspective, they had avoided looking at the main characters, and ignored what they were saying, presumably because the main characters were task-irrelevant. Thus, the effects of adopting the interior decorator perspective, which changed viewer's eye-movements while watching the video clip, seemed to be *at odds* with the default implicit task of understanding the narrative (which was probably much closer to the detective perspective, given that the main characters were discussing a murder).

[6] Indeed, the participants were told that each perspective was defined in terms of tasks, such as, among other things, for the detective perspective, "Your task is to evaluate which persons act suspiciously and could be potential suspects for murder," and for the interior decorator perspective, "Your task is to evaluate how you could improve the interiors and exteriors you see in order to make them more comfortable" (Lahnakoski et al., 2014, p. 324).

We therefore predicted viewer eye-movements in the *Touch of Evil* clip would similarly be affected by a cognitive task that was designed to be specifically at odds with understanding the narrative. We chose a modified version of the spot-the-location task from Smith and Mital (2013). Participants were instructed to "draw a map of the area depicted from memory" after viewing the film clip, "including naming and labeling as many locations as possible" (Full Instructions in Appendix A). We very carefully avoided explicitly instructing viewers where to look, but instead worded the instructions such that it left it up to the viewers to decide how to accomplish the task, including decisions on where to look. Nevertheless, we expected that the task would encourage viewers to look at the background locations (e.g., stores and their signs) in order draw a detailed map with labels. Therefore, as with the Lahnakoski et al. (2014) interior decorator perspective task, we expected that, compared to viewers in the Context condition of the previous experiments, viewers in the Map Task condition would be more likely to look at the periphery of the scene, and make correspondingly longer saccades to do so. We also expected that, compared to viewers in the Context condition, viewers in the Map Task condition would look less at the car, because they would be looking more at the buildings and signs. Likewise, we expected that, because viewers in the Map Task condition would be looking at different things than viewers in the Context condition, we would find reduced gaze similarity between Context condition viewers and Map Task viewers.

Note that all of these predicted effects of higher level cognition on eye-movements while watching the *Touch of Evil* film clip would be evidence of *breaking* the tyranny of film. Nevertheless, we also expected that such effects of the Map Task would not necessarily be overwhelming, and that viewers of the film clip would still be drawn to look at many of the same things at the same time as viewers in Experiments 2 & 4, for the reason that our knowledge of the film clip suggested that some parts of it lacked much background detail to look at (e.g., the medium shot of the couple kissing at the end of the clip shown in Figure 1). Thus, we expected a reduction in the tyranny of film but not a complete elimination of it. Finally, we also predicted that the Map Task would lead to a reduction in understanding the film clip at the level of the event model, with a corresponding decrease in the proportion of participants who would make the critical inference regarding the bomb in the car, because the Map Task viewers' attention would frequently not be on information important to the narrative, but rather on information in the background.

Method

Participants

Data was collected from an additional 75 participants. For this additional experimental condition, a priori criteria were created to select which participants would be included in analyses. First, the Gardony Map Analyzer was used to score participant's maps for accuracy (Gardony, Taylor, & Brunyé, 2015), and only participants with scores greater than or equal to the median were included in any of the analyses (see data analysis section for specifics on the Gardony Map Analyzer. See appendix B for example maps with scores.). Second, only participants that *did not* make the inference about the bomb at the end of the clip were included in the eye-movement analyses. This is because one purpose of the Map Task was to make the narrative irrelevant, and participants able to make the inference about the bomb must have attended to the narrative. This resulted in a total of 37 comparison group participants being included in eye-movement data analyses. Future analyses may look at more fine grained comparisons of all of these participants using map scores as a continuous predictor of eye-movement variability.

Stimuli & procedures

All stimuli and procedures were identical to the Context condition in Experiment 4a except for the inclusion of the Map Task. All participants were presented the Context version of the film clip. Before presentation of the clip, participants were given the Map Task instructions to draw a detailed map of the locations in the scene, including labels, at the end of the clip from memory. After watching the clip, they were prompted to make the inference about what would happen next. After this they were given an 8 1/2'' by 11'' sheet of paper with the instructions printed at the top and grid lines for the map. They had 5 minutes to complete their map.

Data analysis

All predictive inference and eye-movement analyses for the Map Task participants were the same as presented previously. To score the maps, we used the automated Gardony Map Analyzer (Gardony et al., 2015). This starts with a master configuration map given as input to

the software with all relevant locations labeled. Then, each participant's drawn map is scanned, input to the software, its labeled locations are marked, and it is compared to the master map and given a similarity score. We used the SQRT (Canonical Accuracy) measure in the Gardony map analysis program (Gardony et al., 2015), which is a general measure that scores both on the number and configuration of landmarks. To create the master configuration map, Google Earth was used to find the actual streets (in Venice, CA) on which the opening scene of *Touch of Evil* was filmed.[7] With the layout of the street, each of the locations in the clip were placed as accurately as possible to their location on the Google Map. This gave us an objectively accurate map of the scene, which would give participants who drew the most accurate maps the highest scores using the map analyzer.

Results

Predictive inference

Inference data was analyzed for all 75 participants that completed the experiment. Initial coding by two raters showed an interrater reliability of *Kappa* = .945, $p < .001$. Due to the high reliability, the one discrepancy between raters was resolved by discussion to give a single inference made or not made for each participant.

The inference results for the Map Task condition were compared to those in the free viewing Context condition, which we will refer to as the "Comprehension condition" since both conditions viewed the same Context condition video clip, but the free viewing group's implied task was to comprehend the film clip. Participants in the Map Task condition were less likely to make the inference about the bomb (Mean Proportion = .13) than those in the Comprehension condition (Mean Proportion = .50, (X^2 (1, N = 205) = 27.56, $p < .001$, Eta = .367). This indicates that the Map Task was cognitively at odds with the process of narrative comprehension (i.e., all participants had identical visual information available, including seeing the bomb put in the car,

[7] Numerous "Touch of Evil" fan websites discuss the filming location of this classic film, and particularly the famous opening scene in Venice Beach, California. Based on one such site (Robh, 2011), and our own judgment comparing the film with current and archival photos of the area, we chose the streets on Google Maps to use. The shot starts at a lot near the corner of Windward Ave. and Speedway. From there the shot goes South West on Woodward Ave. to what is now Ocean Front Walk. From there the shot goes North West on Ocean Front Walk.

but processed it differently to complete their given tasks). The key question is whether part of the difference in processing was the deployment of overt visual attention as measured by eye-movements.

Eye-movements

Eye-movement analyses compared participants in the Map Task to participants in the Comprehension condition *who made the inference*. These two groups represent participants that most successfully completed their respective tasks; either create a mental map of the scene or comprehend the narrative. Context condition participants that did not make the inference and No-context participants are not presented in the analyses below, because in all previous experiments their eye-movement data did not significantly differ from the Comprehension condition. Similarly, although only the Map Task and Comprehension conditions are reported below, for exploratory purposes, analyses for all other conditions were run. All of the relationships presented hold in ANOVA's that include participants in the Context condition that did not make the inference and in the No-context, as well as participants in the Map Task that did make the inference.

Fixation durations and saccade lengths

All data was cleaned using the same procedures as outlined in Experiment 2. For fixation durations, there were no significant differences between the groups. In the Comprehension group the average fixation duration was 388 ms (SD = 63 ms)[8], and in the Map Task group it was 361 ms (SD = 47 ms)(t (100) = .695, p = .489).

Mean saccade length between groups, however, did show a significant difference. Consistent with our hypotheses based on the results of Lahnakoski et al. (2014) and Smith and Mital (2013), average saccades were longer in the Map Task group (M = 5.39°, SD = .68°) than in the Comprehension group (M = 4.79°, SD = .55°)(t (100) = 4.56, p < .001; d = 0.91). We hypothesized that this would occur because the Map task participants would make longer

[8] Although this is the same data as used in Experiment 4a for the Context + inference group, the average fixation duration presented here is slightly different. This is because the comprehension group average is for the entire video clip rather than just the shared viewing period with the No-context condition analyzed in Experiment 4a.

saccades in order to explore the edges of the scene to complete their task, thereby (at least partially) ignoring the main characters of the narrative that are typically shown near the center of the screen, which would require shorter saccades to explore.

Attentional synchrony

Attentional synchrony results. The gaze similarity analysis compared Comprehension group participants to those in the Map Task across the entirety of the film clip. Figure 11 shows the results of this comparison.

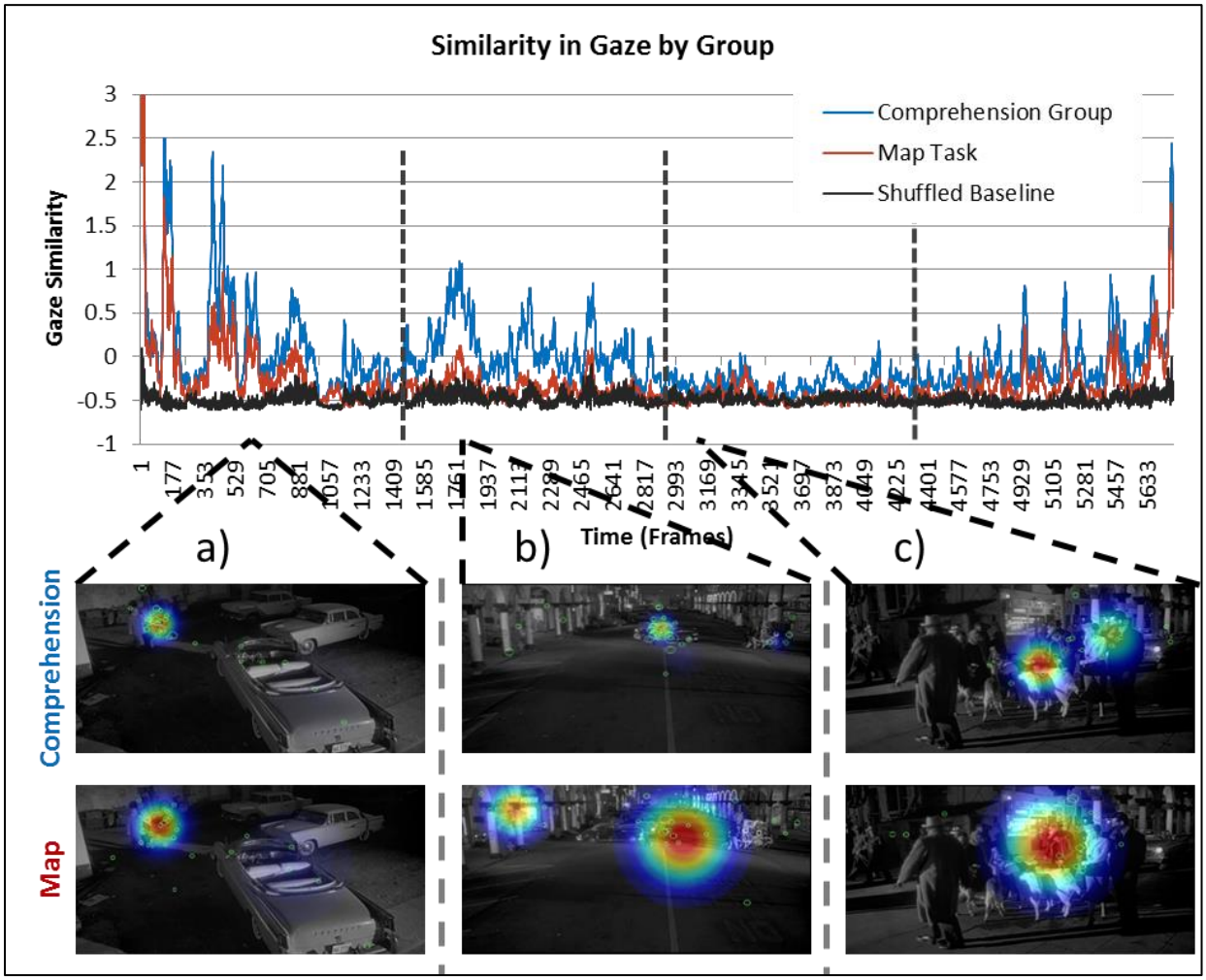


Figure 11. Top: Similarity of gaze by context condition across the full clip. Gaze similarity is expressed as a z-score probability relative to the context condition and inference made (Comprehension group [Blue], Map Task [Red], and Shuffled Baseline [Black]). Large values indicate greater attentional synchrony. Vertical dashed grey lines illustrate the quarters used in the repeated measures analysis. Bottom: The stills exemplify gaze patterns during a) high gaze

similarity (GS) for both groups, b) high GS for Comprehension, low for Map Task, and c) low GS for both.

Qualitatively, Figure 11 shows that gaze similarity scores for the Comprehension group are often higher than for the Map Task group, indicating that viewers in the Map Task were looking at different places on those given frames than the Comprehension group. To quantify this relationship an ANOVA of mean gaze similarity by group was calculated. Consistent with the qualitative assessment of the figure, an ANOVA supports the difference between each group (F (2, 166) = 96.484, $p < .001$, $\eta_{p2} = .541$), with Bonferroni corrected pairwise comparisons showing the comprehension group had the greatest gaze similarity ($M = .001$, $SD = .266$), followed by the Map Task ($M = -.284$, $SD = .234$), and the shuffled baseline ($M = -.488$, $SD = .045$) was the lowest. Nevertheless, Figure 11 also shows that the Map Task group was frequently above the shuffled baseline, and mimicked many of the peaks and troughs of the Comprehension group. This supports the prediction that even when the task is at odds with comprehension, it may be difficult to completely ignore areas associated with comprehension. The tyranny of film is perhaps not being turned off, but it is being turned down.

In addition to the main effects, an inspection of Figure 11 indicates that there may be an interaction of gaze similarity with time. An exploratory, repeated measures ANOVA of task and time in the clip was conducted to better understand the role the features of the clip play in determining gaze similarity. As noted in the Introduction, we chose the *Touch of Evil* film clip, in part, because it has bottom-up features that should reduce attentional synchrony. However, the presence of visual features that may guide attention changes throughout the clip. For the repeated measures ANOVA, the clip was broken into quarters that correspond well to the changes in visual features in the clip (more detail on these features in the interpretation below). The sphericity assumption was violated (X^2 (5) = 36.077, $p < .001$), thus a Greenhouse-Geisser correction was used ($\mathcal{E} = .869$). As with the one-way ANOVA, there was a main effect of time block (F (2.60, 427.76) = 177.113, $p < .001$, $\eta_{p2} = .519$). Bonferroni corrected pairwise comparisons showed that each time block was significantly different from the others; block 1 had the highest gaze similarity ($M = -.114$, $SD = .419$), followed by block 2 ($M = -.227$, $SD = .355$), then block 4 ($M = -.266$, $SD = .313$), and block 3 had the lowest gaze similarity ($M = -.406$, $SD = .151$). Additionally, the task by time block interaction was significant (F (5.217, 427.758) =

71.430, $p < .001$, $\eta_{p2} = .466$). The interaction was probed using simple effects where time block was held constant over task. The Greenhouse-Geisser corrected omnibus error term was *MSE* = .016, and the *df* = 427.758. Gaze similarity differed between all tasks in blocks 1, 2, and 4 (*p*'s < .001, *F*'s > 20.312)[9]. However, in block 3, the Map task and the shuffled baseline were not significantly different F (1, 427.758) = 2.375, $p > .05$, while all other task comparisons in block 3 were significantly different (*p*'s < .001, *F*'s > 34.187).

The main effect of time block and the interaction with task support previous work on the features that guide attentional synchrony when a qualitative analysis of the features of a film clip in each block is used. As shown in Figure 11 (Top) Block 1 has the highest gaze similarity, and qualitatively has features that would support this (Figure 11, Bottom, A), (e.g., close ups of the bomb and car, and relatively little else to look at). Figure 11 shows that Blocks 2 and 4 which have moderate levels of gaze similarity and moderately more to look at (Figure 11, Bottom, B), (e.g., more store fronts). Block 3 has the lowest gaze similarity, and the most complex composition involving lots of people, vehicles, animals, and store fronts (Figure 11, Bottom, C). The complexity in Block 3 is precisely what we predicted would reduce attentional synchrony overall when we chose the *Touch of Evil* film clip (Cutting et al., 2011; Dorr et al., 2010; Mital et al., 2010). Additionally, the large number of storefronts and spread out locations in Block 3 is what would be predicted to reduce gaze similarity in the Map Task, consistent with the lack of a difference between the Map Task and the shuffled baseline in Block 3.

Region of interest

Region of interest results. The same region of interest data pre-processing as in Experiment 2 was used. The car was identified as the region of interest, and we tested whether participants in the Comprehension group fixated the car more often than in the Map Task group. The region of interest analysis was carried out over the entire viewing period.

As can be seen in Figure 12 the red line for participants that completed the Map Task is fairly consistently below that of the blue line for the Comprehension group, which is not surprising considering that car fixations vary with the size of the car on the screen. A *t*-test comparing the proportion of fixations of the car (when on the screen) for each group confirmed

[9] The full simple effects structure is in Appendix B

that participants in the Comprehension group fixated the car significantly more often (10.2% of fixations) than participants in the Map Task group (6.5% of fixations)(t (100) = 3.706, p < .001; d = 0.74). A similar result was shown when overall dwell time on the car was calculated for each group. As with the proportion of fixations, the total time spent on the car was larger for participants in the Comprehension group (16.9 sec; SD = 10.8 sec) than for the Map group (11.6 sec, SD = 7.1 sec)(t (100) = 2.653, p = .009; d = 0.53).

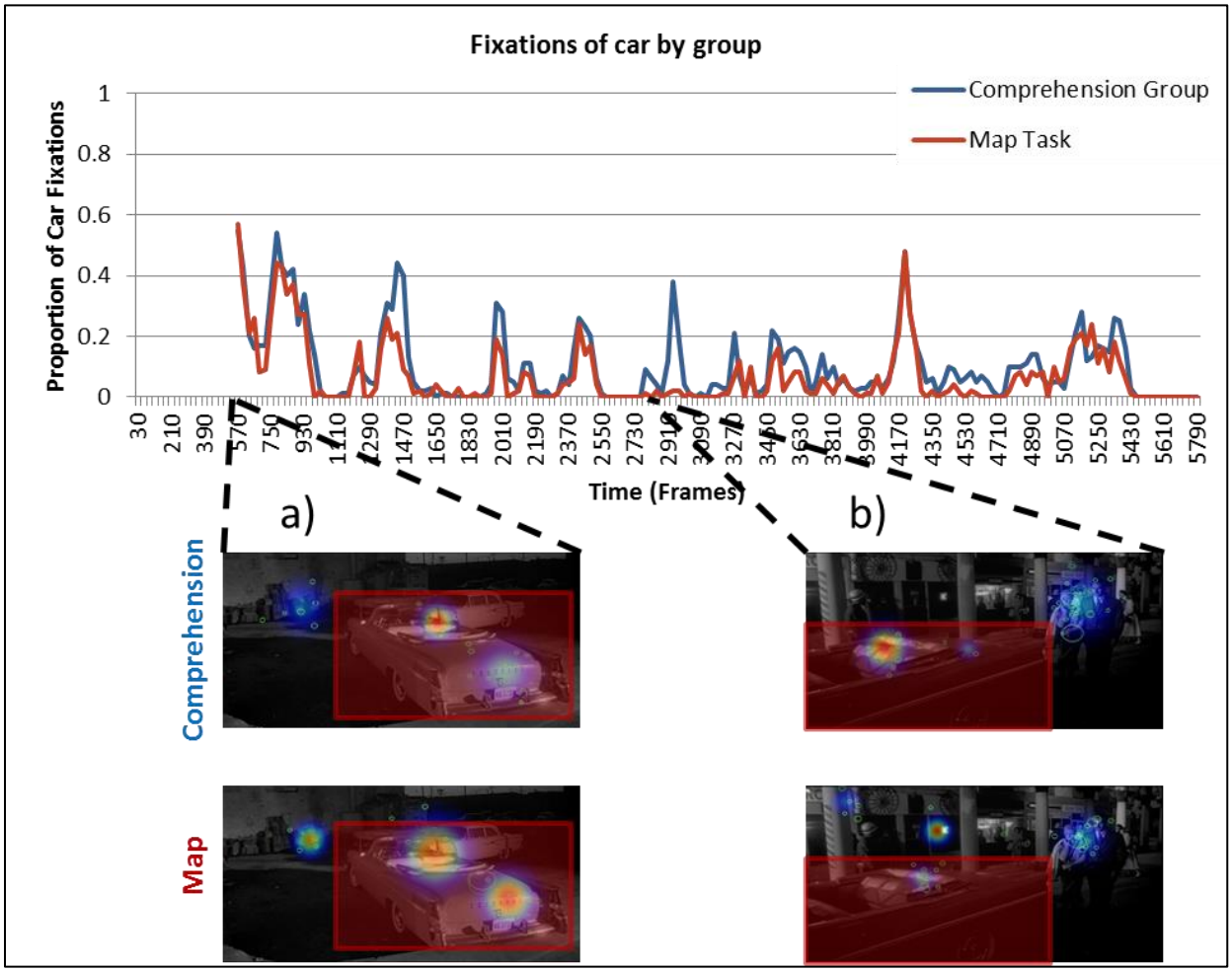


Figure 12. Top: Proportion of participants fixating the car throughout the film clip (Comprehension Group [Blue], Map Task Group [Red]). The higher the value the more participants looked at the car. The car first appears in the clip at frame 541, which is in time bin 570. Bottom: Film stills show the region of interest for the frame, with fixation and heat maps superimposed. The stills indicate a) when both groups fixated the car at a similar rate, and b) when the car first reappears and was highly fixated by the Comprehension group, but not the Map group.

Taken together, the region of interest results are consistent with our predictions for the Map Task. The car is integral for comprehension of the film clip, but is relatively unimportant for completing the Map Task. Accordingly, participants in the Comprehension group look at the car more than participants in the Map Task. As a result, the region of interest results again show the tyranny of film being reduced through the high level cognitive task manipulation of the Map Task. Nevertheless, as with the gaze similarity analysis above, participants in the Map Task condition still looked at the car occasionally, indicating that the reduction of the tyranny of film was relative, but not complete.

Discussion

Experiment 4a tested the role of context on comprehension and eye-movements, and Experiment 4b tested the role of task. The results of 4a showed a limited effect of comprehension, as manipulated by context, on eye-movements. That limited effect was only for fixation durations and during a critical period when there was a difference in the perceived agents in the film, which affected visual attention. To test whether the tyranny of film could be *turned-off* by an explicit task at odds with comprehending the film narrative, Experiment 4b compared the Context + inference group with the Map Task condition (and since comprehending the narrative was the implicit task for the participants in the Context + inference group, we relabeled it as the *Comprehension condition* for that comparison). The results of this comparison showed a meaningful and statistically significant reduction in both the gaze similarity and probability of looking at the chief area of interest (the car) due to the task manipulation. Thus, we can say that an explicit viewing task that was at odds with the task of comprehension *turned-down* the tyranny of film. Because we showed experimentally induced variation in the degree of the tyranny of film, this suggests that our previously shown dissociation between comprehension (as manipulated by context) and eye-movements may be a *true null effect* rather than being due to a weak manipulation of comprehension or poor measures of the cognitive effects on eye-movements. This is something that will need to be tested more, as the likelihood that it is a true null effect can be ascertained through a programmatic approach of attempts to reject the null. It also suggests that the online control of eye-movements in dynamic scenes is highly task dependent.

Chapter 6 - General Discussion

When we watch visual narratives in film, television, and perhaps in our real world environment we seem to be able to comprehend them fairly easily. How is it that we might be able to take purely visual input and create a coherent narrative of the events that unfold? Through a series of four experiments we found strong effects of knowledge of the bomb on comprehension on predictive inference generation and event segmentation. Despite this, there was an overall lack of an effect on eye-movements, except when a cognitive task at odds with narrative comprehension was used.

Rethinking the tyranny of film

One potential reason for the overall lack of an effect of comprehension on eye-movements with *Touch of Evil* (Welles, 1958) could be that even though it was chosen for what appeared to be weak bottom-up features, it still guided participant eye-movements as much as the *Moonraker* (Broccoli & Gilbert, 1979) film clip used in the previous study (Loschky et al., 2015). However, Figures 13a and 13b clearly show *Moonraker* has more clustering on fewer clusters than *Touch of Evil*. Furthermore, as would be expected based on the gaze similarity results, the Map Task had even less clustering than all other conditions[10]. Thus, it seems that even films with relatively weak bottom-up features show little effect of comprehension on eye-movements. *Touch of Evil* does have relatively weak bottom-up features when compared to *Moonraker*, yet the comprehension and eye-movement results are analogous.

[10] The lack of a difference in the number of clusters for the Map Task and the other *Touch of Evil* conditions makes sense, because as the overall clustering lessens there are going to be fewer groupings of gaze that are identified as a cluster.

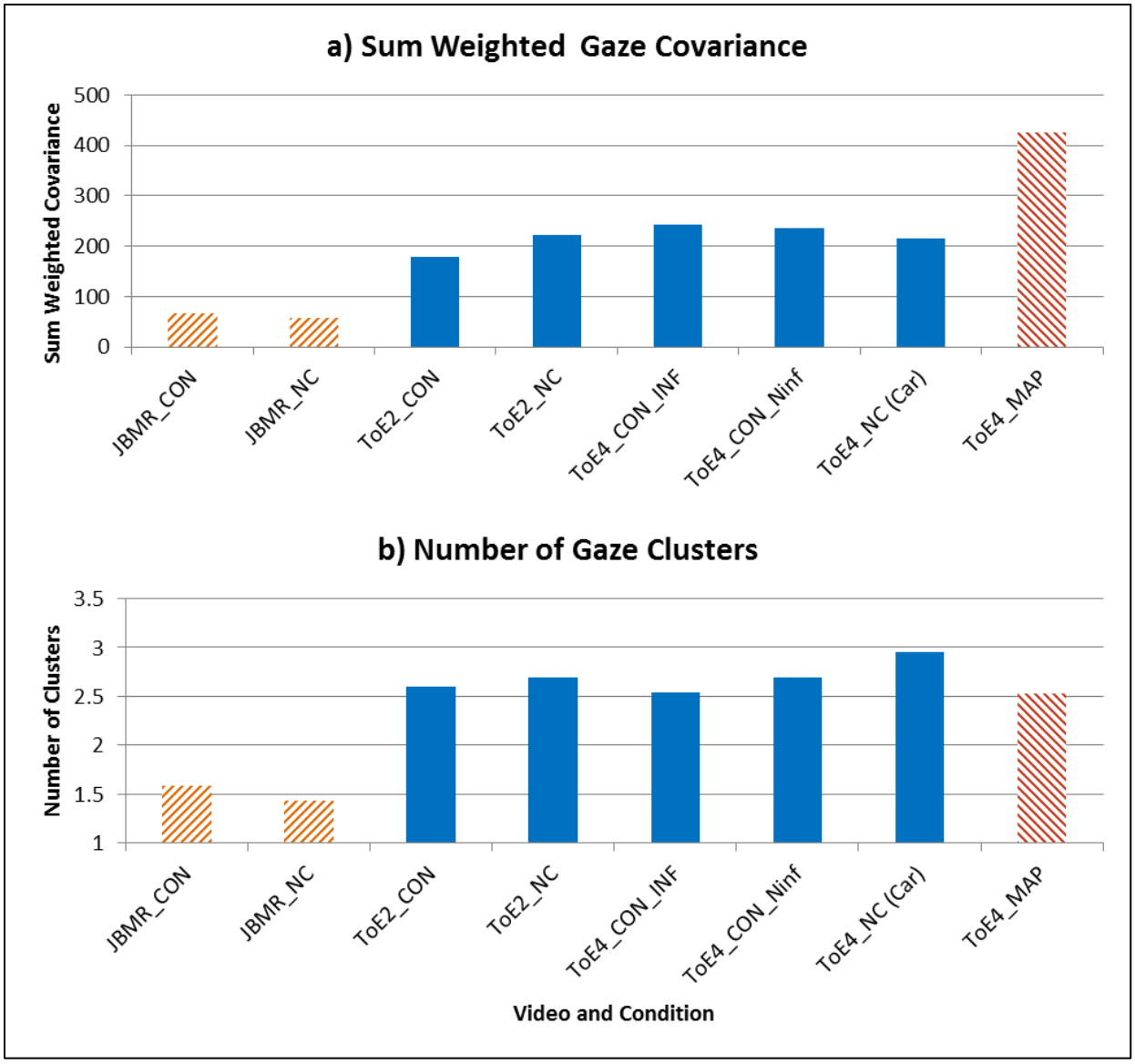


Figure 13. a) The average sum weighted gaze covariance for all eye-tracking conditions in the James Bond *Moonraker* (Loschky et al., 2015) and *Touch of Evil* studies. The orange bars are from the *Moonraker* study, the blue bars are from the *Touch of Evil* comprehension conditions, and the red bar is the *Touch of Evil Map task*. b) The average number of gaze clusters for the same experiments.

These results call for a reconsideration of the tyranny of film hypothesis, which states that there is no opportunity for differences in comprehension to be expressed through differences in eye-movements due to bottom-up guidance of viewer attention by visual features. Clearly, fewer people are looking at the same places at the same times in *Touch of Evil* than in

Moonraker, yet, we still find few if any effects of large differences in comprehension (as manipulated by context) on eye-movements.

Top-down attention is slow and effortful. One reason for the lack of a top-down effect during film viewing may be the relationship between film comprehension and eye-movements is simply very weak at best. Eye-movement models offer some support for the weak relationship between bottom-up and top-down processes during film viewing. For example, highly produced films cut frequently, creating a new visual scene every few seconds which may keep the visual system in a stage of early processing that relies more heavily on bottom-up features (Mital et al, 2010). The *Touch of Evil* clip was chosen in part because it doesn't have editing that introduces a new shot every 2-3 seconds, but it may be that the introduction of new information due to the camera continuously tracking in the shot creates a similar effect of keeping the viewer in an early stage of processing that relies more on bottom up attention.

Within the Findlay and Walker (1999) model of saccade generation, there are early stages of processing that are said to be automatic initially, then automated, and then more volitional. What this means is that in early visual processing of a scene saccades are carried out before the visual system can voluntarily select information. In the current study, the voluntary selection would be of bomb-relevant information in the narrative. The automatic and automated processing could occur in *Touch of Evil* because eye-movements are being driven only by bottom-up Behaviorally this is consistent with work showing that it takes approximately 2-3 seconds before eye-movements show strong effects of top-down processes in scenes (Parkhurst, Law, & Niebur, 2002; Tatler, Baddeley, & Gilchrist, 2005; Zelinsky, Adeli, & Vitu, 2016).

The reason for slower top-down effects in scenes is that executive control of eye-movements is effortful and uses more executive attentional resources. Such executive control of eye-movements generally involves the activation of the frontal eye fields (FEF), which are essential for performing the anti-saccade task (i.e., looking in the opposite direction an equal distance from a target that appears to the left or right of fixation) (Findlay & Walker, 1999; Guitton, Buchtel, & Douglas, 1985; Odriscoll et al., 1995; Sweeney, Mintun, Kwee, & Wiseman, 1996). Working memory resources are also influential in explicit cognitive control, and such tasks become more difficult as one's memory resources decline due to dual tasking, aging, or brain damage (Mitchell, Macrae, & Gilchrist, 2002; Roberts, Hager, & Heron, 1994).

Conversely, the pro-saccade task (i.e., looking at a target that appears to the left or right of fixation) and the capture of attention by salient stimuli strongly involves the superior colliculus and the posterior parietal cortex (Boehnke & Munoz, 2008; Kustov & Robinson, 1996, November 7). We hypothesize that this seemingly effortless control of eye-movements in the pro-saccade task may be more similar to watching a film for comprehension. Further, we hypothesize there is a fundamental difference between the control of eye-movements during the comprehension of film, and task based control of eye-movements for the Map Task that required participants to inhibit processing of the focal features of the clip in order to deploy attention to the periphery of the shot. Further research is needed to directly test this hypothesis. This could be done by testing the effect of cognitive loads on eye-movements during the performance of comprehension versus others tasks such as the Map Task. If the Map Task requires more executive control than the comprehension task, a cognitive load would be expected to more greatly affect eye-movements during the Map Task than the comprehension task (i.e., participants in the Map Task would have eye-movements more similar to those in the Comprehension condition).

When comprehension has an effect. Although top-down processes are slow and effortful, there was the agent effect in Experiment 4, and small effects of the event model were also found in Loschky et al. (2015). Thus, film comprehension can have an impact on visual attention during film viewing. Interestingly, the effects found may have occurred through different mechanisms. The agent effect found in the current study appears to have occurred because participants in the No-context condition continued to track the agents (the walking couple) in their event model, which may have been a more automated process using the term from Findlay and Walker (1999). Conversely, the effect in Loschky et al. (2015) appears to have occurred due to the No-context group needing to use effortful process to update their mental model with what was presented. Based on this, differences in eye-movements may be related to the amount of effortful processing the viewer used to comprehend a scene. The Map Task similarly required effortful updating to successfully complete the task. The importance of the potential need for effortful processing is that it may indicate the variable strength of the different components of mental models used to comprehend narratives (Zwaan, Langston, & Graesser, 1995), and how effortful their updating is. In the four experiments, we only manipulated two of

the five indices of the event indexing model (agent and causality). A more comprehensive study manipulating all five indices (agent, time, place, causality, and intentionality [goal-relevance]) may give a clearer picture of the effect of each of these event indices on attention. We hypothesize that when an event index must be effortfully updated to maintain comprehension for a narrative, it will predictably guide eye-movements.

More generally, the different effect sizes of the Map Task and the agent effect point toward a continuum of higher level cognitive effects on eye-movements during narrative film viewing. The Map Task is more similar to the task manipulations that have been used to fairly consistently show strong top-down effects on eye-movements in scenes and videos (DeAngelus & Pelz, 2009; Lahnakoski et al., 2014; Smith & Mital, 2013; Yarbus, 1967). Conversely, the agent effect may be a much weaker top-down process on the continuum, but is still strong enough to break the tyranny of film. An even weaker effect, or overall lack of an effect, may be the null effect of expertise found in Taya et al. (2012).

There is much work to be done on the larger continuum of top-down effects of higher level cognition on eye-movements (for review, Baluch & Itti, 2011). For example, in real life, viewers continuously selectively expose themselves to the media they are interested in by choosing the websites they go to, the movies they attend, and the television channels they flip to (Hart et al., 2009; Knobloch-Westerwick & Meng, 2009; Stroud, 2008). On a more micro level, movie viewers sometimes close their eyes during certain parts of movies, or avert their gaze from something they do not want to see in the detail afforded by the fovea. Similarly, if a film viewer is asked to only look at the top right corner of the screen, this is something that most could do, although depending on their interest in the film they are watching, it could be a somewhat difficult, unenjoyable task. Therefore, tasks, such as the Map Task, that are at odds with the filmmaker's objectives, will require viewers to exert executive control of their overt attention (Fan, McCandliss, Fossella, Flombaum, & Posner, 2005; Fan, McCandliss, Sommer, Raz, & Posner, 2002).

What breaks the tyranny of film? Despite the ambiguity of what drives the tyranny of film, the dissociation between eye-movements and narrative comprehension is very surprising in *Touch of Evil*, especially with the strength and consistency of the differences in comprehension between the groups. These findings are inconsistent with the majority of previous work looking

at top-down task-based effects on scene viewing (Foulsham & Underwood, 2007; Henderson et al., 2007; Henderson et al., 2013; Smith & Mital, 2013; Yarbus, 1967).

One characteristic of this finding is that it is inconsistent with a strict interpretation of the "eye-mind hypothesis" (Just & Carpenter, 1980; Reichle et al., 1998; Reilly & Radach, 2006). It appears that the dimension of causality, in relation to the bomb, in participants' mental models is not guiding eye-movements, while agency is having a small effect. This does not necessarily indicate that these components cannot affect eye-movements. It is possible that, for example, because the bomb is hidden in the trunk of the car that viewers chose not to look at its hiding place.[11] This interpretation of the results would argue for a weaker version of the "eye-mind hypothesis." Namely, depending on viewers' tasks and goals, there can be dissociations between eye-movements and thought (Lamont, Henderson, & Smith, 2010; Smith, 2015).

This weak version of the "eye-mind hypothesis" may be due in part to film viewing being driven by both bottom-up features and *mandatory* top-down processes (Baluch & Itti, 2011). Mandatory top-down processes are well-learned, more automated processes, as opposed to volitional top-down processes. A classic mandatory process is the hollow face illusion, when a concave face (e.g., the inside of a mask), is perceived as convex (Baluch & Itti, 2011; Gregory, 1970). Another example is following the speaker of a conversation (Birmingham et al., 2008). Within film viewing, the task of comprehension may have certain mandatory processes used to construct a mental model. One of these could be to identify and locate key agents in the narrative, and attend to them, as suggested by the agent effect found in Experiment 4a. This type of top-down process may operate at a lower level than the manipulation causality used in the current study.

One recent theory that may support this is the role of the default mode network in narrative comprehension (Tylén et al., 2015). That is, the default mode network may allow for the accumulation of coherent plot information. Conversely, when plot information is less coherent, the frontoparietal control network is thought to allow for a more effortful search for narrative coherence through a more top-down deployment of attention. The reason for this is

[11] Of course, our results suggest that about half of the viewers who knew about the bomb were unable to hold it in their WM to the end of the clip. Yet, those participants (Context+No-inference) were just as likely to look at the car as those who did not forget it (Context+Inference), or those who never knew about the bomb (No-context+No-inference).

that the default mode network has been shown to be less active when visual attention is effortfully deployed (Andrews-Hanna, Reidler, Huang, & Buckner, 2010; Andrews-Hanna, Reidler, Sepulcre, Poulin, & Buckner, 2010). When considering the highly complex visual stimulus of a film, it may seem that the default mode network should not play a large role, but fMRI research with film viewing has indicated activation of the default mode network (Hasson et al., 2008; Hasson, Malach, & Heeger, 2010). Additionally, areas thought to make up part of the default mode network have been shown to be similarly activated during both film viewing and audio book listening (Hasson et al., 2008).

Within film viewing, breaking the tyranny of film may have two potential paths. The first could be to directly tap into a mandatory process (e.g., agent tracking). If the narrative one viewer perceives in a scene has entirely different characters than the narrative another viewer perceives in the same scene, they should track different agents. This was one of the hypotheses tested in the current study, but the film clip used appears to have been well constructed to give high importance to both the walking couple and the couple in the car. Specifically, the car is initially tracked by the camera, but when the walking couple is introduced the camera begins to track them while the car with the bomb lurks more in the periphery of the frame, or off-screen. The other track to breaking the tyranny of film is to move away from mandatory processing and automated comprehension processes. The Map Task appears to have done this, but it should be possible with a comprehension manipulation as well. For example, in Loschky et al. (2015), the effect on eye-movements occurred during a complex cross cutting sequence that required viewers to make an inference (that both sequences would come together in time and space and solve the life-and-death problem faced by the protagonist in one of the two sequences). The viewers that had more trouble making the inference about a critical shot showed eye-movement differences during that shot. However, this shot was essentially a static scene, and thus did not use any film features to guide viewer attention. Nevertheless, future work could test if a break in coherence allows viewers to move from mandatory processing to more effortful, volitional processing even during dynamic scenes in a film.

Applications. Dynamic stimuli such as film are often designed to guide eye-movements based on the purpose of viewing the dynamic stimulus (e.g., film narrative comprehension). However, there are many situations during which it may be beneficial for a viewer to attend

somewhere other than the location specified by the dynamic stimulus. One common example of this is the use of videos in a classroom setting. Instructors will often use videos as examples, even though the video may have been designed for another purpose. For example, a Women's Studies instructor may want to guide students' attention to aspects of the film other than those intended by the film-maker (e.g., to gender-stereotyped elements in a scene which are simply considered by the film-maker to a part of the scene's background). The difficulty of breaking the tyranny of film with the Map Task indicates that instructors may have a similarly difficult time guiding their students' attention to other appropriate information in a video clip that was not specifically designed for the instructor's pedagogical purposes.

A poor control of attention during video viewing could have detrimental effects when it comes to learning. Research on eye-movements and problem solving has shown a strong connection between where people look and their ability to solve a problem (Grant & Spivey, 2003; Madsen, Larson, Loschky, & Rebello, 2012). Additionally, and more importantly for the current research, cues that guide eye-movements to the appropriate areas can increase the probability of correct problem solving (Madsen, Rouinfar, Larson, Loschky, & Rebello, 2013; Rouinfar, Agra, Larson, Rebello, & Loschky, 2014; Thomas & Lleras, 2007). This establishes a strong, bidirectional relationship between comprehension and eye-movements in problem solving. Within film, when a viewer has little ability to volitionally control their eye-movements, guidance to the wrong areas could lead to an inability to attend to the appropriate information, potentially an incorrect understanding of concepts, and hinder critical thinking on the topic.

Summary

The current study tested whether a person's comprehension during film viewing affects their eye-movements. The differences in comprehension we found were consistent with similar research in reading comprehension, but novel to film comprehension research. However, despite these large comprehension differences, similar to our previous study that used a very different film clip (Loschky et al., 2015) we found only small and targeted differences in eye-movements. These findings are counterintuitive based on work looking at top-down effects on eye-movements in static scenes (Smith & Mital, 2013; Yarbus, 1967), but consistent with the finding of strong attentional synchrony in film viewing (Dorr et al., 2010; Mital et al., 2010; Smith &

Mital, 2013; Wang et al., 2012). Based on this, the tyranny of film hypothesis was mostly supported. The comprehension processes used in visual narratives seem to be similar to those used in reading. This is based on the similarities between the relationship of working memory with comprehension and the event segmentation results in this study with what has been shown in reading studies (Allbritton, 2004; Calvo, 2001, 2005; Daneman & Carpenter, 1980; Daneman & Merikle, 1996; Just & Carpenter, 1992; Kurby & Zacks, 2008; Linderholm, 2002; Rai et al., 2014; Rai et al., 2011; Speer & Zacks, 2005; St George et al., 1997; Zacks et al., 2009). However, the processes by which information is extracted through eye-movements appears to differ between visual narratives and reading (Magliano et al., 2013), as seen in the general dissociation between eye-movements and comprehension. The results are interesting in terms of both film comprehension processes and eye-movement processes in film perception, but the dissociation of these processes may be the most interesting. During typical film viewing people may attend to the same places, but have different understandings of the narrative. This is a counter-example to the common assumption in many eye-movement studies that there is a strong association between eye-movements and thought (Just & Carpenter, 1980; Reichle et al., 1998; Reilly & Radach, 2006). The inclusion of the Map Task condition in Experiment 4 indicates that tasks at odds with film narrative comprehension can provide support for cognitive control of eye-movements, supporting the eye-mind hypothesis. However, at the level of mental model construction during film narratives, the underlying task may be too similar to allow for large eye-movement differences to be expressed. To better understand this dissociation of eye-movements and comprehension during film viewing, future studies need to combine techniques and theories from the fields of scene perception, event perception, and narrative comprehension (Loschky, Hutson, Magliano, Larson, & Smith, 2016; Loschky, Hutson, Magliano, Larson, & Smith, 2014, June; Magliano, Larson, Higgs, & Loschky, 2016). Enriching our understanding of when and how we perceive and comprehend visual information from our environment can lead to a better understanding of even higher order processes such as decision making, reasoning (Madsen et al., 2013; McNamara & Magliano, 2009; Rouinfar et al., 2014; Thomas & Lleras, 2007), and generally how we interact with our immediate environment, because comprehension of our environment is integral to our interaction with (Loschky et al., 2016; Loschky et al., 2014, June; Magliano et al., 2016).

References

Allbritton, D. (2004). Strategic production of predictive inferences during comprehension. *Discourse Processes, 38*(3), 309-322.

Andrews-Hanna, J. R., Reidler, J. S., Huang, C., & Buckner, R. L. (2010). Evidence for the default network's role in spontaneous cognition. *Journal Of Neurophysiology, 104*(1), 322-335.

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., & Buckner, R. L. (2010). Functional-anatomic fractionation of the brain's default network. *Neuron, 65*(4), 550-562.

Bach, M. (2006). The Freiburg Visual Acuity Test-Variability unchanged by post-hoc re-analysis. *Graefe's Archive for Clinical and Experimental Ophthalmology, 245*(7), 965-971. doi:10.1007/s00417-006-0474-4

Baluch, F., & Itti, L. (2011). Mechanisms of top-down attention. *Trends in Neurosciences, 34*(4), 210-224. doi:http://dx.doi.org/10.1016/j.tins.2011.02.003

Bezdek, M. A., Gerrig, R. J., Wenzel, W. G., Shin, J., Revill, K. P., & Schumacher, E. H. (2015). Neural evidence that suspense narrows attentional focus. *Neuroscience, 303*, 338-345.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Gaze selection in complex social scenes. *Visual Cognition, 16*(2-3), 341-355. doi:10.1080/13506280701434532

Boehnke, S. E., & Munoz, D. P. (2008). On the importance of the transient visual response in the superior colliculus. *Current Opinion In Neurobiology, 18*(6), 544-551. doi:10.1016/j.conb.2008.11.004

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 717-726. doi:http://dx.doi.org/10.1016/S0022-5371(72)80006-9

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology, 13*(2), 207-230.

Broccoli, A. R. P., & Gilbert, L. D. (Writers). (1979). Moonraker [Film]: Available from CBS/Fox Video, Industrial Park Drive, Farmington Hills, MI 48024.

Calvo, M. G. (2001). Working memory and inferences: Evidence from eye fixations during reading. *Memory, 9*(4-6), 365-381.

Calvo, M. G. (2005). Relative contribution of vocabulary knowledge and working memory span to elaborative inferences in reading. *Learning And Individual Differences, 15*(1), 53-65.

Carmi, R., & Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research, 46*(26), 4333-4345. doi:http://dx.doi.org/10.1016/j.visres.2006.08.019

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*(3), 386-404. doi:http://dx.doi.org/10.1016/0022-0965(82)90054-6

Coutrot, A., & Guyader, N. (2014). How saliency, faces, and sound influence gaze in dynamic social scenes. *Journal of Vision, 14*(8), 5.

Cutting, J. E. (2015). The framing of characters in popular movies. *Art & Perception, 3*(2), 191-212.

Cutting, J. E., Brunick, K. L., DeLong, J. E., Iricinschi, C., & Candan, A. (2011). Quicker, faster, darker: Changes in Hollywood film over 75 years. *i-Perception, 2*(6), 569-576.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466.

Daneman, M., & Merikle, P. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*(4), 422-433. doi:10.3758/bf03214546

DeAngelus, M., & Pelz, J. (2009). Top-down control of eye movements: Yarbus revisited. *Visual Cognition, 17*(6), 790 - 811.

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision, 10*(10). doi:10.1167/10.10.28

Eisenberg, M. L., & Zacks, J. M. (2016). Ambient and focal visual processing of naturalistic activity. *Journal of Vision, 16*(2), 5-5.

Fan, J., McCandliss, B., Fossella, J., Flombaum, J., & Posner, M. I. (2005). The activation of attentional networks. *Neuroimage, 26*(2), 471-479.

Fan, J., McCandliss, B., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of cognitive neuroscience, 14*(3), 340-347.

Findlay, J., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences, 22*(4), 661-721.

Foulsham, T., & Underwood, G. (2007). How does the purpose of inspection influence the potency of visual salience in scene perception? *Perception, 36*(8), 1123-1138.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*(2), 178-210.

Gardony, A. L., Taylor, H. A., & Brunyé, T. T. (2015). Gardony Map Drawing Analyzer: Software for quantitative analysis of sketch maps. *Behavior Research Methods*, 1-27.

Gernsbacher, M. A. (1990). *Language comprehension as structure building* (Vol. xi). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science, 14*(5), 462-466. doi:10.1111/1467-9280.02454

Gregory, R. L. (1970). *The intelligent eye* (1st ed.). New York, NY: McGraw-Hill.

Guitton, D., Buchtel, H. A., & Douglas, R. M. (1985). Frontal lobe lesions in man cause difficulties in supressing reflexive glances and in generating goal directed saccades. *Experimental brain research, 58,* 455-472.

Hard, B. M., Tversky, B., & Lang, D. S. (2006). Making sense of abstract events: Building event schemas. *Memory & Cognition, 34*(6), 1221-1235. doi:10.3758/bf03193267

Hart, W., Albarracín, D., Eagly, A., Brechan, I., Lindberg, M., & Merrill, L. (2009). Feeling validated versus being correct: a meta-analysis of selective exposure to information. *Psychological Bulletin, 135*(4), 555.

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., & Heeger, D. J. (2008). Neurocinematics: The neuroscience of film. *Projections, 2*(1), 1-26.

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences, 14*(1), 40-48. doi:http://dx.doi.org/10.1016/j.tics.2009.10.011

Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science, 16*(4), 219-222.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. L. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. v. Gompel, M. Fischer, W. Murray, & R. W. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537-562). Amsterdam: Elsevier.

Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (Vol. xi, pp. 269-293). Oxford, England: Elsevier Science Ltd.

Henderson, J. M., Shinkareva, S. V., Wang, J., Luke, S. G., & Olejarczyk, J. (2013). Predicting cognitive state from eye movements. *PLoS ONE, 8*(5), e64937. doi:10.1371/journal.pone.0064937

Ho, S., Foulsham, T., & Kingstone, A. (2015). Speaking and listening with the eyes: gaze signaling during dyadic interactions. *PLoS ONE, 10*(8), e0136905.

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition, 12*(6), 1093-1123.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*: Harvard University Press.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329-354.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122-149.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press

Knobloch-Westerwick, S., & Meng, J. (2009). Looking the other way selective exposure to attitude-consistent and counterattitudinal political information. *Communication Research, 36*(3), 426-448.

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences, 12*(2), 72.

Kustov, A. A., & Robinson, D. L. (1996, November 7). Shared neural control of attentional shifts and eye movements. *Nature, 384*(6604), 74-77.

Lahnakoski, J. M., Glerean, E., Jääskeläinen, I. P., Hyönä, J., Hari, R., Sams, M., & Nummenmaa, L. (2014). Synchronous brain activity across individuals underlies shared psychological perspectives. *Neuroimage, 100*, 316-324.

Lamont, P., Henderson, J. M., & Smith, T. J. (2010). Where science and magic meet: The illusion of a ‚Äúscience of magic‚Äù. *Review of General Psychology, 14*(1), 16-21.

Latif, N., Gehmacher, A., Castelhano, M. S., & Munhall, K. G. (2014). The art of gaze guidance. *Journal of Experimental Psychology: Human Perception and Performance, 40*(1), 33.

Linderholm, T. (2002). Predictive inference generation as a function of working memory capacity and causal text constraints. *Discourse Processes, 34*(3), 259-280.

Loschky, L. C., Hutson, J. P., Magliano, J. P., Larson, A., & Smith, T. (2016). *The Scene Perception and Event Comprehension Theory (SPECT) Applied to Visual Narratives*. Paper presented at the International Conference on Empirical Studies of Literature and Media, Chicago, IL.

Loschky, L. C., Hutson, J. P., Magliano, J. P., Larson, A. M., & Smith, T. J. (2014, June). *Explaining the Film Comprehension/Attention Relationship with the Scene Perception and Event Comprehension Theory (SPECT)*. Paper presented at the The 2014 annual meeting of the Society for Cognitive Studies of the Moving Image, Lancaster, PA.

Loschky, L. C., Larson, A. M., Magliano, J. P., & Smith, T. J. (2015). What Would Jaws Do? The Tyranny of Film and the Relationship between Gaze and Higher-Level Narrative Film Comprehension. *PLoS ONE, 10*(11), e0142474. doi:10.1371/journal.pone.0142474

Loughlin, S. M., & Alexander, P. A. (2012). Explicating and exemplifying empiricist and cognitivist paradigms in the study of human learning *Paradigms in theory construction* (pp. 273-296): Springer.

Madsen, A. M., Larson, A. M., Loschky, L. C., & Rebello, N. S. (2012). Differences in visual attention between those who correctly and incorrectly answer physics problems. *Physical Review Special Topics - Physics Education Research, 8*(1), 010122-010121-010113.

Madsen, A. M., Rouinfar, A., Larson, A. M., Loschky, L. C., & Rebello, N. S. (2013). Can short duration visual cues influence students' reasoning and eye movements in physics problems? *Physical Review Special Topics - Physics Education Research, 9*(2), 020104-020101 -020104-020116.

Magliano, J. P., Larson, A. M., Higgs, K., & Loschky, L. C. (2015). The relative roles of visuospatial and linguistic working memory systems in generating inferences during visual narrative comprehension. *Memory & Cognition*, 1-13.

Magliano, J. P., Larson, A. M., Higgs, K., & Loschky, L. C. (2016). The relative roles of visuospatial and linguistic working memory systems in generating inferences during visual narrative comprehension. *Memory & Cognition, 44*(2), 207-219.

Magliano, J. P., Loschky, L. C., Clinton, J., & Larson, A. M. (2013). Is reading the same as viewing? An exploration of the similarities and differences between processing text- and visually based narratives. In B. Miller, L. Cutting, & P. McCardle (Eds.), *Unraveling the Behavioral, Neurobiological, and Genetic Components of Reading Comprehension* (pp. 78-90). Baltimore, MD: Brookes Publishing Co.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1997). Fixation sequences made during visual examination of briefly presented 2D images. *Spatial Vision, 11*(2), 157-178.

McNamara, D. S., & Magliano, J. P. (2009). Toward a comprehensive model of comprehension. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. Volume 51, pp. 297-384): Academic Press.

Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2010). Clustering of Gaze During Dynamic Scene Viewing is Predicted by Motion. *Cognitive Computation, 3*(1), 5-24. doi:10.1007/s12559-010-9074-z

Mitchell, J. P., Macrae, C. N., & Gilchrist, I. D. (2002). Working Memory and the Suppression of Reflexive Saccades. *Journal of cognitive neuroscience, 14*(1), 95-103. doi:10.1162/089892902317205357

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal Of Personality And Social Psychology, 28*(1), 28.

Odriscoll, G. A., Alpert, N. M., Matthysse, S. W., Levy, D. L., Rauch, S. L., & Holzman, P. S. (1995). Functional Neuroanatomy of Antisaccade Eye Movements Investigated With Positron Emission Tomography. *Proceedings of the National Academy of Sciences of the United States of America, 92*(3), 925-929.

Pannasch, S., Helmert, J. R., Roth, K., Herbold, A. K., & Walter, H. (2008). Visual fixation durations and saccade amplitudes: Shifting relationship in a variety of conditions. *Journal of Eye Movement Research, 2*(2:4), 1-19.

Parkhurst, D. J., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42*(1), 107-123.

Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research, 45*(18), 2397-2416.

Rai, M. K., Loschky, L. C., & Harris, R. J. (2014). The effects of stress on reading: A comparison of first language versus intermediate second-language reading comprehension. *Journal of Educational Psychology*. doi:http://dx.doi.org/10.1037/a0037591

Rai, M. K., Loschky, L. C., Harris, R. J., Peck, N. R., & Cook, L. G. (2011). Effects of stress and working memory capacity on foreign language readers' inferential processing during comprehension. *Language learning, 61*(1), 187-218. doi:10.1111/j.1467-9922.2010.00592.x

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372-422.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*(1), 125-157.

Reilly, R. G., & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research, 7*(1), 34-55. doi:http://dx.doi.org/10.1016/j.cogsys.2005.07.006

Reingold, E. M., & Stampe, D. M. (2000). Saccadic inhibition and gaze contingent research paradigms. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 119-145). Amsterdam: Elsevier.

Reingold, E. M., & Stampe, D. M. (2002). Saccadic inhibition in voluntary and reflexive saccades. *Journal of cognitive neuroscience, 14*(3), 371-388.

Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive Science, 31*(4), 613-643.

Roberts, R. J., Hager, L. D., & Heron, C. (1994). Prefrontal cognitive processes: Working memory and inhibition in the antisaccade task. *Journal of Experimental Psychology: General, 123*(4), Dec 1994, 1374-1393.

Robh. (2011, January 28, 2011). Touch of Evil opening shot location. Retrieved from http://broadviewgraphics.blogspot.com/2011/01/touch-of-evil-opening-shot-location.html

Rouinfar, A., Agra, E., Larson, A. M., Rebello, N. S., & Loschky, L. C. (2014). Linking attentional processes and conceptual problem solving: Visual cues facilitate the automaticity of extracting relevant information from diagrams. *Frontiers in psychology, 5*. doi:10.3389/fpsyg.2014.01094

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception, 28*(9), 1059-1074.

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review, 5*(4), 644-649.

Smith, T. J. (2012). The attentional theory of cinematic continuity. *Projections, 6*(1), 1-27. doi:10.3167/proj.2012.060102

Smith, T. J. (2013). Watching you watch movies: Using eye tracking to inform cognitive film theory. In A. P. Shimamura (Ed.), *Psychocinematics: Exploring Cognition at the Movies*. New York: Oxford University Press.

Smith, T. J. (2015). The role of audience participation and task relevance on change detection during a card trick. *Frontiers in psychology, 6*.

Smith, T. J., Lamont, P., & Henderson, J. M. (2012). The penny drops: Change blindness at fixation. *Perception, 41*(4), 489-492.

Smith, T. J., Levin, D. T., & Cutting, J. E. (2012). A window on reality: Perceiving edited moving images. *Current Directions in Psychological Science, 21*(2), 107-113. doi:10.1177/0963721412437407

Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behaviour in static and dynamic scenes. *Journal of Vision, 13*(8), 16.

Speer, N. K., Swallow, K. M., & Zacks, J. M. (2003). Activation of human motion processing areas during event perception. *Cognitive and Affective Behavioral Neuroscience, 3*(4), 335-345.

Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. *Journal of memory and language, 53*(1), 125-140.

St George, M., Mannes, S., & Hoffman, J. E. (1997). Individual differences in inference generation: An ERP analysis. *Journal of cognitive neuroscience, 9*(6), 776-787.

Stroud, N. J. (2008). Media use and political predispositions: Revisiting the concept of selective exposure. *Political Behavior, 30*(3), 341-366.

Sweeney, J. A., Mintun, M. A., Kwee, S., & Wiseman, M. B. (1996). Positron emission tomography study of voluntary saccadic eye movements and spatial working memory. *Journal Of Neurophysiology, 75*(1), 454-468.

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research, 45*(5), 643-659.

Taya, S., Windridge, D., & Osman, M. (2012). Looking to score: The dissociation of goal influence on eye movement and meta-attentional allocation in a complex dynamic natural scene. *PLoS ONE, 7*(6), 1-9.

Thomas, L. E., & Lleras, A. (2007). Moving eyes and moving thought: On the spatial compatibility between eye movements and cognition. *Psychonomic Bulletin & Review, 14*(4), 663-668.

Tully, T. (1999). The sounds of evil. *Videography Magazine, January*.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of memory and language, 28*(2), 127-154. doi:http://dx.doi.org/10.1016/0749-596X(89)90040-5

Tylén, K., Christensen, P., Roepstorff, A., Lund, T., Østergaard, S., & Donald, M. (2015). Brains striving for coherence: Long-term cumulative plot formation in the default mode network. *Neuroimage, 121*, 106-114.

van Dijk, T., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Vo, M. L. H., Smith, T. J., Mital, P. K., & Henderson, J. M. (2012). Do the eyes really have it. *Dynamic allocation of.*

Wang, H. X., Freeman, J., Merriam, E. P., Hasson, U., & Heeger, D. J. (2012). Temporal eye movement strategies during naturalistic viewing. *Journal of Vision, 12*(1). doi:10.1167/12.1.16

Welles, O. (Writer). (1958). Touch of Evil [film]: Universal Pictures.

Yarbus, A. L. (1967). *Eye movements and vision* (B. Haigh, Trans.). New York, NY, US: Plenum Press.

Zacks, J. M. (1999). *Event Structure Perception Studies in Perceiving, Remembering, and Communicating*: Stanford University.

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General, 138*(2), 307-327. doi:2009-05547-010 [pii] 10.1037/a0015305

Zacks, J. M., Speer, N. K., Swallow, K. M., & Maley, C. J. (2010). The brain's cutting-room floor: Segmentation of narrative cinema. *Frontiers in Human Neuroscience, 4*. doi:10.3389/fnhum.2010.00168

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127*(1), 3-21.

Zelinsky, G., Adeli, H., & Vitu, F. (2016). *The new best model of visual search can be found in the brain.* Paper presented at the Vision Sciences Society, St. Pete Beach, FL.

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. doi:10.1111/j.1467-9280.1995.tb00513.x

Appendix A - Map Task Instructions

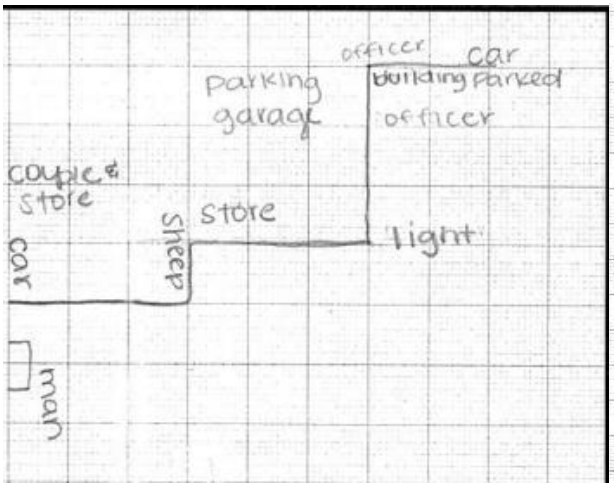
Map Task Instructions:

*Your task is to **watch a video** clip of a town, and after you finish, **draw a map** of the area depicted from memory. Your map should be as detailed as possible **including naming and labeling** as many locations as possible. Your map will be scored for its level of detail and accuracy. You have 5 minutes to complete your map.*

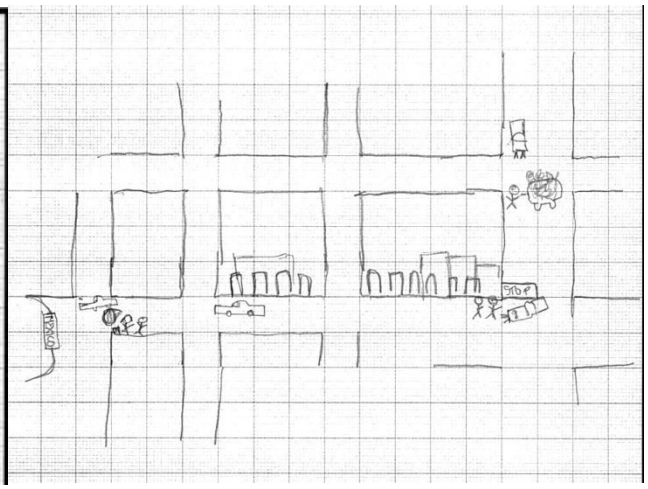
Appendix B - Map Task Score Examples

Low Map Score (0.0)

a) Ambiguous Labels

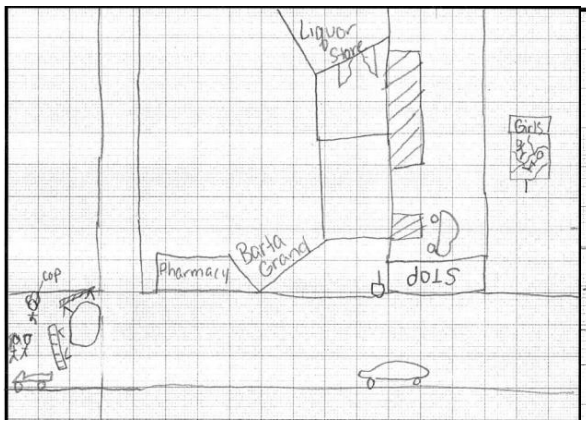


b) Locations not identified

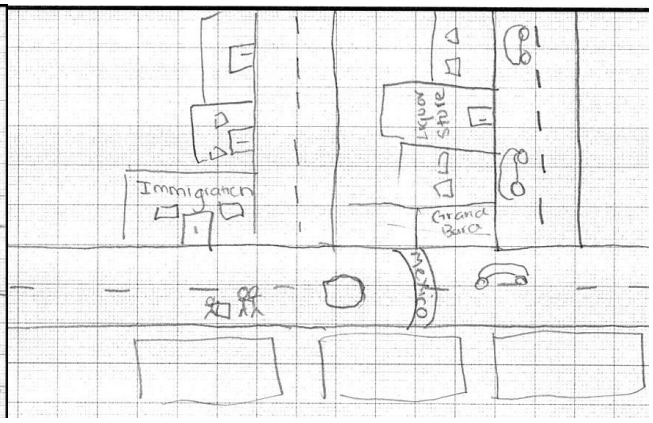


Median Map Score (.17)

a) 4 locations labeled

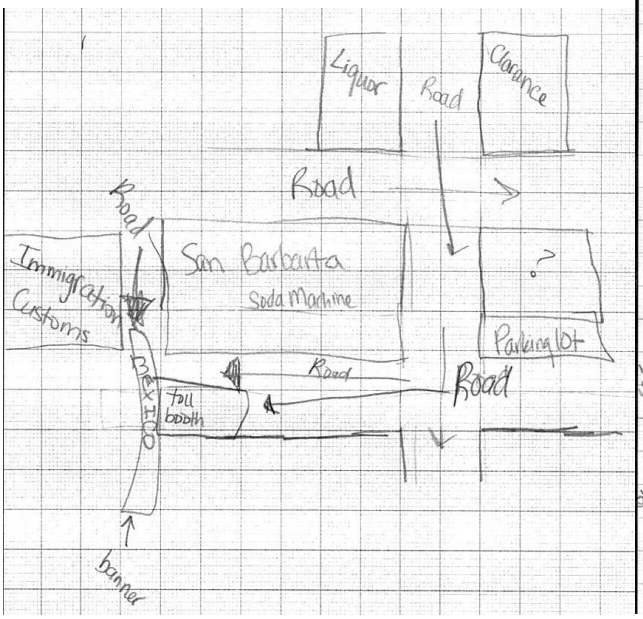


b) 4 locations labeled

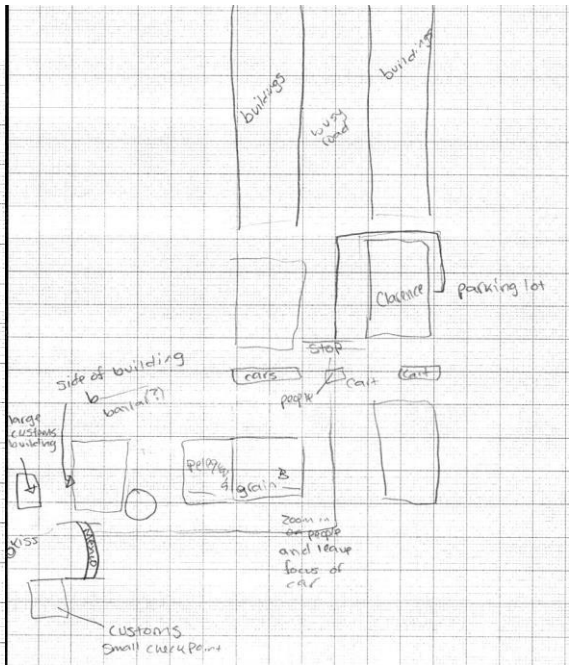


Top-Scores (.37 & .38)

a) .37 (7 locations given near correct location)



b) .38 (7 locations given near correct location)



Appendix C - Repeated Measures Analysis of Variance Simple Effects

Simple Effects:

Quarter held constant over condition:

The omnibus error term was .016 and the degrees of freedom were 427.758.

Quarter 1:

- Context and Map: $F = 2.417/.016$; F (1, 427.758) = 151.062, p <.001
- Context and Shuffle $F = 18.020/.016$; F (1, 427.758) = 1126.25, p <.001
- Map and Shuffle $F = 4.248/.016$; F (1, 427.758) = 265.5, p <.001

Quarter 2:

- Context and Map: $F = 4.426/.016$; F (1, 427.758) = 276.625, p <.001
- Context and Shuffle $F = 9.857/.016$; F (1, 427.758) = 616.062, p <.001
- Map and Shuffle $F = .325/.016$; F (1, 427.758) = 20.312, p <.001

Quarter 3:

- Context and Map: $F = .547/.016$; F (1, 427.758) = 34.187, p <.001
- Context and Shuffle $F = 1.204/.016$; F (1, 427.758) = 75.25, p <.001
- **Map and Shuffle $F = .038/.016$; F (1, 427.758) = 2.375, $p > .05$**

Quarter 4:

- Context and Map: $F = 1.254/.016$; F (1, 427.758) = 78.375, p <.001
- Context and Shuffle $F = 7.059/.016$; F (1, 427.758) = 441.188, p <.001
- Map and Shuffle $F = 1.307/.016$; F (1, 427.758) = 81.688, p <.001