How do individual differences impact attentional selection ability during volitional control tasks in film?

by

Taylor Leigh Simonson

B.A., The University of Findlay, 2017

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

2021

Approved by:

Major Professor
Dr. Lester Loschky

Abstract

Film viewers' eye-movements seem largely disconnected from their comprehension; viewers' eye-movements rarely deviate from narrative elements, regardless of differences in their comprehension (Loschky et al., 2015; Hutson et al., 2017), suggesting bottom-up film features (e.g., color or motion) overwhelm top-down attentional control (e.g., knowledge or experiences). However, viewers' eye-movements were shown to deviate from narrative elements when given a task at odds with comprehension (Hutson et al., 2017). This suggests viewers used volitional attentional control, which has been suggested to be cognitively demanding during tasks like the anti-saccade task. In such cases, participants tend to make more incorrect, reflexive eye-movements (Mitchell et al., 2002), suggesting difficultly (e.g., cognitive demand) in volitionally controlling their eye-movements. Thus, can people readily use volitional attention during film viewing or is it cognitively demanding for them do so?

Furthermore, there have been conflicting results surrounding eye-movement differences across Western and Eastern cultures (Nisbett & Masuda, 2003; Rayner, Castlehano, & Yang, 2007; Goh, Tan, & Park, 2009). Thus, will these cultural differences be evident in our data when using film stimuli, rather than static images, or will bottom-up film features be too overwhelming to show such cultural differences?

Finally, studies have shown an important role of working memory capacity in attentional control during cognitively demanding tasks (Unsworth, Schrock, & Engle, 2004; Unsworth & Engle, 2007). Further, these studies have shown differences in performance impact attention control as well. Thus, what role will individual differences (e.g., working memory or task performance) have in attentional selection ability, if any?

Participants from Kansas and Japan were eye-tracked while viewing film clips. During film viewing, they were presented with different task goals and levels of attentional demand. Specifically, participants had a primary task of either watching a film clip for comprehension (Comprehension Condition) or drawing a map of the film space from memory (Map Condition). Participants had a secondary task (cognitive load) on half the trials to increase attentional demand. After their working memory capacity was assessed

Generally, results showed clearer cultural effects than in past studies using static images, showing that cultural background can overwhelm the otherwise controlling bottom-up features in film. There results also replicated and extended past studies using volitional control measures during film viewing, showing changes in film viewers' eye-movements depending on the tasks they were given. Importantly, such volitional control of film viewers' eye-movements came at a cost to film comprehension and was shown to be cognitively demanding. Additionally, there were effects that suggested performance was changing over time on our different performance measures, that working memory played a role in performance, and some evidence to suggest task trade-offs were occurring for Kansas and not Kyoto.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to thank my master's committee members: Dr. Lester Loschky who guided me through every aspect of this project from start to finish; Dr. Heather Bailey for her unique insights into the cognitive load and working memory aspects of Chapter 2 specifically; and Dr. Jun Saiki who provided enormous support in the cultural comparison areas of this project, and all other aspects of it, and graciously hosting me for one month in Kyoto to work on the study there in his lab. I would also like to thank my other Kyoto University Collaborators, Dr. Yoshiyuki Ueda, Yu Yana, and Shunsuke for helping with all matters related to the cross cultural portion of this study, and Ryoh Takamori for helping implement the experiment in Kyoto; Dr. John Hutson for helping in early stages of project creation to get things running as well as throughout the study on various analyses and data related questions; Ashley Wiegel and Jazmin Royg-Quevedo on scoring the Map data in Kansas; Yu Aoki, Nao Mitsubayashi, Takato Yamazaki, Yoshizumi Yamazaki for scoring map data in Kyoto.; Yuhang Ma, Ella McCloud, Kenzi Kriss, Anna Cook, & Katherine Kolze for their work on the narrative-AOIs creation; the members (past and present) of the Visual Cognition Lab at Kansas State University for their continued support on all aspects of this project; and Lisa Vangsness, Maverick Smith, and Destiny Bell for their continued suggestions on the paper while writing!

Chapter 1 - The Role of a Goal and Culture on Eye-Movements: Volitional and Mandatory Top-Down Attentional Selection During Film Viewing

Attention

Viewers pay attention to what they care about. Attention is often measured with eye-movements, which serve as a measure for what we attend to, how long we attend, and in what order our environment draws our attention in a scene (Henderson, 2007). Eye-movements have allowed for the discovery of two types of influences on attentional selection that occur when viewers are shown static or dynamic scenes (Henderson, 2007; Henderson & Hollingworth, 1998): bottom-up (e.g., stimulus features like color or motion) and top-down attentional selection influences (e.g., viewer influences like experiences or goals). These two influences of attentional selection have been independently researched in great depth and have allowed researchers to more accurately and fully characterize how viewers attend to and interact with information in their environments (Yarbus, 1967; DeAngelus & Pelz, 2009; Dorr, Martinez, Gegenfurtner, & Barth, 2010).

Bottom-up and top-down influences on attention can be found in both lab settings that do not approximate the real world well (e.g., static image presentations; Yarbus, 1967; Henderson, 2007; DeAngelus & Pelz, 2009) and viewing conditions that are more naturalistic and approximate the real world (e.g., dynamic image presentations; Dorr et al., 2010; Smith and Mital, 2013). Research on static scenes (e.g., photographs) and dynamic scenes (e.g., film clips) have provided a more complete understanding of how these processes influence our attention in everyday life. To this point, attentional selection studies have largely utilized static images rather than dynamic images, which would be more like real-life.

Bottom-up influences on attentional selection

Bottom-up attention is driven by the properties of the stimulus (e.g., more salient features like bright color, motion, or contrast; Henderson, 2007). More specifically, participants automatically switch their attention to a suddenly appearing stimulus even when they are told not to do so (Mitchell, Macrae, & Gilchrist, 2002; Unsworth, Schrock, & Engle, 2004; Unsworth & Engle, 2007). Thus, attentional selection may be an automatic process that is not impacted by our knowledge for information in the stimulus we are shown.

Top-down influences on attentional selection

Top-down processes are viewer driven, meaning that the knowledge, memories, and experiences of the viewer have an impact on what they attend to (Henderson, 2007). For example, eye-movement patterns depend on the task that participants are given (Yarbus, 1967; DeAngelus & Pelz, 2009). When Yarbus (1967), and later DeAngelus and Pelz (2009), asked viewers to identify the ages of the people in a picture, eye-movements focused on faces more than furniture; but when viewers were asked to identify the socioeconomic status of the people in the image, they focused on the furniture and clothing instead. This demonstrates that the knowledge and experiences that we have for the world influence the attentional selection patterns that we engage in, which are top-down processes at work.

Recent studies have moved away from using static images to utilizing dynamic images, which are more analogous to real-life situations. In contrast to the work that has been done with static images, work with dynamic images has shown that it is difficult to find evidence for top-down effects on attention because people tend to look at the same places at the same times. One source of this effect is the use of "cuts" that directors/editors often use during films to encourage viewers to look in the same places at the same times (Smith & Henderson, 2008; Mital et al.,

2011), a phenomenon known as *attentional synchrony* (Dorr, et al., 2010; Smith, Levin, & Cutting, 2012; Smith & Mital, 2013).

Dorr and colleagues (2010) showed that when comparing different types of stimuli (e.g., static scenes, stop motion clips, and dynamic scenes with and without cuts) that the strongest central bias, and thus attentional synchrony, occurred when viewing Hollywood style film trailers, which used filmmaking techniques. There was a central bias with other types of films, such as more naturalistic clips that did not have any filmmaking techniques present, but nothing that was as strong as those in the Hollywood style film clips. This suggests when studying attentional selection in film, longer shots with little-to-no-editing should be used to reduce more artificially produced attentional synchrony. This has been done by some researchers by using unedited, long shots from Hollywood-style films, rather than clips or trailers that have edits and cuts throughout (Hutson, Smith, Magliano, & Loschky, 2017). This specific type of Hollywood-style film, which has less attentional synchrony, is closer to a naturalistic viewing state, which makes the results more applicable to real-life attentional selection behaviors. This has led to a growing body of literature that investigates the ways in which top-down processes are used to view dynamic images in these more naturalistic dynamic scenes (Carmi & Itti, 2006; Smith & Mital, 2013; Lahnakoski et al., 2014; Loschky, Larson, Magliano, & Smith, 2015; Hutson, et al., 2017).

Types of top-down attention**.** Researchers who are interested in the role of top-down processes have made the distinction between two types of top-down attentional selection processes: mandatory and volitional processes (Baluch & Itti, 2011). Mandatory top-down effects are more automatic and are influenced by a person's prior experiences and knowledge (2011), or even cultural background (Karden, et al., 2017). For example, it is normal for attention

to focus on the faces of people; this could be the result of a lifetime of experience of looking at people's faces, which are critically important for daily interactions with others. Conversely, volitional attention is the act of willfully shifting your attention to something. Volitional effects of attention are seen when someone is given a goal, such as when a student is in class and must decide what to-be-learned information is most valuable to reaching and completing their learning goals. The influence of volitional attention within an academic setting helps to highlight the importance of studying volitional top-down attention and its impact on a person. Understanding how volitional attention and comprehension interact in this context is critical to continue the study of attention within film to model real-life behaviors.

The tyranny of film. A study was done to investigate how knowledge that participants have for a film clip impacts their eye-movements (Loschky et al., 2015), since attentional selection while watching films have not been shown to be impacted by top-down processes. More specifically, Loschky et al. (2015) looked at the role of comprehension on attentional selection. Their hypothesis was that manipulating viewers' comprehension would produce differences in eye-movements between the two conditions. To do this, the study utilized the jumped in the middle paradigm, which is the laboratory equivalent to walking in on the middle of a television show and having no background information on the narrative before that moment. Specifically, in one condition participants saw the full film 3-minute clip and, in another condition, participants saw only the last 12 seconds of the film clip, so they had no context for the events they saw. However, despite differences in comprehension between the two conditions, there were only limited differences in the viewers' eye-movements. This suggests filmmakers use cinematic techniques to create a strong bottom-up influence on viewers' attentional selection during typical Hollywood style films. This phenomenon was coined the Tyranny of Film

(Loschky et al., 2015), which is a lack of differences in viewers' eye-movements despite large differences in their comprehension.

Volitional control of attention. Follow-up studies have tested the limits of this concept and to "break the tyranny." Hutson et al. (2017) manipulated context to change how participants interpreted the main narrative elements during a film clip from Orsen Welles's Touch of Evil (Welles & Zugsmith, Touch of Evil, 1958). Specifically, the context condition witnessed a bomb being placed in the trunk of a car, and then a couple unknowingly got into the car and drove away. In contrast, the no-context condition did not see the portion of the film clip showing the bomb. Hutson et al. (2017) hypothesized that by eliminating knowledge of the bomb, participants' understanding of the clip in the two conditions would be radically different, particularly concerning the relevance of the car. This would differentially influence participants' attention to the car while watching the clip. Shockingly, it also supported the Tyranny of Film hypothesis, by finding large differences in comprehension did not result in large differences in attentional selection during film viewing.

Hutson et al. (2017) further manipulated the context by having participants in the no-context condition start watching the clip even later, when only a walking couple were shown on the screen, in order to manipulate who participants perceived as the main characters in the clip. This produced a predicted, but small and limited difference in attention between the groups. Specifically, when the no-context condition later saw the car for the first time, they paid much less attention to it than those in the context condition. This is assumedly because the no-context condition participants did not consider the couple in the car to be main characters. However, this only lasted for the first eight seconds where the car was seen by the no-context group. Thereafter,

they looked at the car just as much as participants in the context condition, thus only small and limited effects were found.

Importantly, in another condition, Hutson et al. (2017) changed the goal of the participants, by giving participants a task to complete, which was at odds with comprehending the narrative of the film clip. This task was to watch the film clip in preparation to draw a map of the objects and locations in the film space from memory once the film clip was over. This *map task* required participants to use volitional control of their attention, by shifting their attention away from the main narrative elements of the film clip. Participants who completed the map task had more varied eye-movements, meaning less attentional synchrony, thus attenuating the Tyranny of Film. The results showed that participants' attention was drawn away from the narrative and moved to the background elements in the periphery of the screen, because of the volitional attentional control map task participants were given.

Importantly, participants in the map task condition showed evidence of having impaired event models for the narrative, as only 10% made a predictive inference that the bomb would explode, compared to 50% of participants in the context condition. This showed that the map task was indeed at odds with comprehending the narrative, assumedly because it drew participants' attention away from the central narrative elements (e.g., the main characters), which, in turn, attenuated the tyranny of film. Nevertheless, the result that a volitional task at odds with comprehending the narrative attenuated the tyranny of film, while interesting, is limited in one critical way. Mainly, the result came from a single film clip, and so needs to be replicated with a larger set of film clips before considering it as being generalizable.

Cultural Comparisons

Cultural background has been a debated topic within the attention literature as to whether it plays a large role in attentional selection or not. Differences have been observed between western (e.g., individualistic cultures like the US, Europe, and Canada) and eastern cultures (e.g., collectivist cultures like Japan, China, and Korea; Masuda & Nisbett, 2001; Nisbett & Masuda, 2003; Chua, Boland, & Nisbett, 2005). These studies show that Western cultures focus on stimuli in a more analytic way, meaning that the salient foreground objects are more commonly the focus of their attention. Conversely, Eastern cultures often focus on stimuli in a more holistic way, with attention set on the background more often, rather than the salient foreground objects. Ultimately, this results in Eastern cultures remembering more about how the background and foreground relate to one another (2003). These results have been found across a variety of measures such as attention, perception, memory, and comprehension.

Alternatively, studies have found no evidence of differences between cultures and rather have shown that the two cultures behave quite similarly when viewing images. Specifically, they found no, or small, differences in fixation durations (Evans et al., 2009; Rayner et al., 2009); no differences in the number of total fixations (Evans et al., 2009; Goh, Tan, & Park, 2009; Rayner et al., 2009; Masuda, Ishii, & Kimura, 2016) or the number of fixations to background and central elements (Chua, Boland, & Nisbett, 2005; Evans et al., 2009; Goh, Tan, & Park, 2009; Masuda, Ishii, & Kimura, 2016). These studies largely show that both cultures tend to prefer the central objects of an image to an equal degree, while Eastern participants may spend more time than western cultures exploring the background space but not by a large degree. More recently it seems that the literature is leaning towards no differences in attentional selection between different cultures. This leaves the current literature to be quite divided. An interesting point is

that most culture studies focus on the use of static images, rather than dynamic images, which begs the question of whether this type of mandatory top-down attentional process would have an influence over attentional selection processes or not. However, it is possible that there would be too many strong bottom-up features, such as motion, for such attentional differences to still be present.

Overview of the Current Study

The aim of this study was trifold. Firstly, we wanted to replicate and extend the finding that the tyranny of film can be overridden while viewing film with a task at odds with comprehending the film narrative (Hutson, et al., 2017). Secondly, we wanted to identify if the hotly debated cross-cultural differences in attentional selection between Western European cultures and East Asian cultures would be apparent while watching film (Masuda & Nisbett, 2001; Nisbett & Masuda, 2003; Chua, Boland, & Nisbett, 2005; Rayner, 2007; Goh, Tan, & Park, 2009; Evans et al., 2009; Rayner, Castelhano, & Yang, 2009) or if attention synchrony while viewing film would be too strong for culture to influence attentional selection, namely whether the cultural influences on attention would be subject to the tyranny of film. Thirdly, we wanted to identify if volitional control of attention is cognitively demanding (this is not the primary focus of this paper and can be further explored in the supplementary materials)[1].

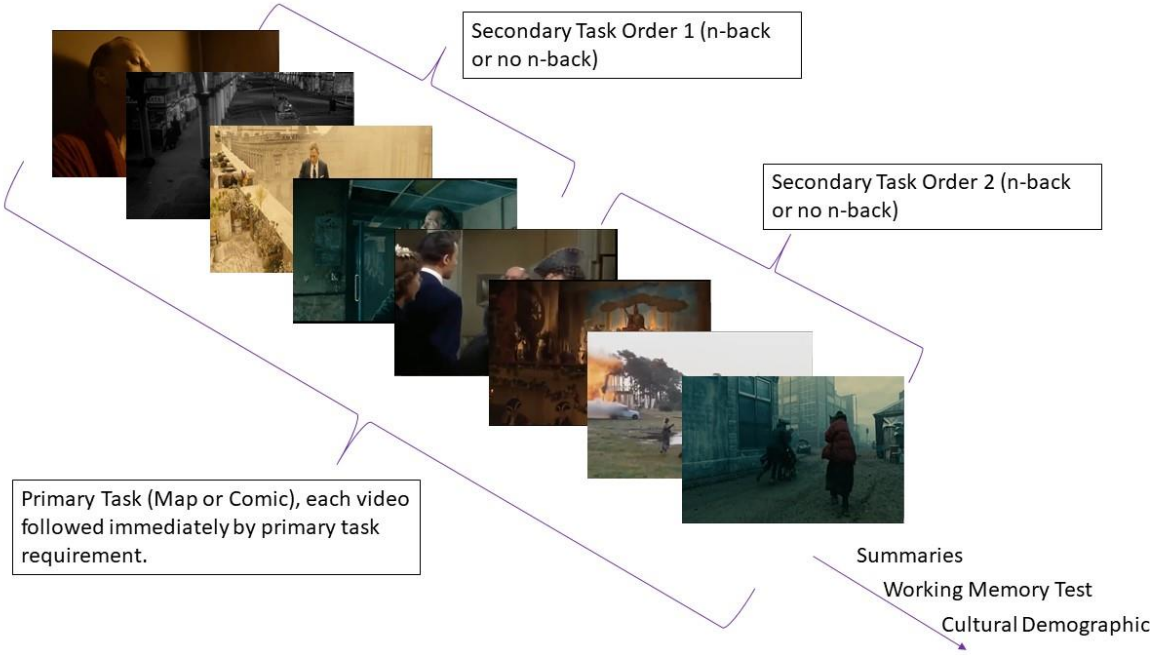
The experimental schematic of the study is depicted in Figure 1.1 below. This study utilized 8 short film clips that participants watched while preparing for 1 of 2 primary tasks: the

1. [1] The variable of cognitive load had no effect on any main outcome measures and was thus not included in the main body of the paper, assumedly because the cognitive load created by the auditory 1-Back task was so small as to be negligible. Instead, those results are reported in the Supplementary Materials. This was done to focus the paper on the meaningful results while also simplifying the description of design of the study and making the study easier for readers to follow.

map task (i.e., draw a map of the landmarks (or objects) and their locations in the film space from memory after watching the film) or the comic task (i.e., draw a 4-panel comic depicting the main narrative events in the film from memory after watching the film). These primary tasks served to manipulate participants' goals while watching the film clips to manipulate their volitional attention. Additionally, on half of the trials (in either the first or second block of 4 videos), participants completed a secondary cognitive load task (see supplementary materials for more information on this). After watching all videos, participants wrote summaries for all the film clips, completed an executive working memory measure, and then filled out a cultural demographic questionnaire.

Figure 1.1.

Trial Schematic for Current Study



Notes. Experiment schematic for stimuli, tasks, and conditions. 8 video clips counterbalanced across participants. The primary task (map or comic task condition) was completed after each video and was varied between-subjects. The secondary task (cognitive load or none) was done while watching videos, either during the first or second block of four videos, was varied within-subjects and counterbalanced. After all videos, participants wrote summaries for each of the videos, took an executive working memory test.

Hypotheses

The *volitional attention hypothesis* predicts that participants given a goal at odds with comprehension of a narrative will have significantly different eye-movements compared to participants who were not given a task at odds with comprehension. The *Cultural Dependence hypothesis* predicts that the cultural background of a person will influence their attentional guidance during film viewing.

While these two hypotheses are not competing, they share a single ***competing alternative*** hypothesis. The *Tyranny of Film hypothesis* predicts that despite large differences in cognition

(e.g., task or cultural background) while watching the film, there will be little or no differences in eye-movements because of the strong bottom-up features of the film[2].

Method

Participants

There was a total of 86 participants, with participants recruited from Kansas State University (Kansas; n = 46; 31 Females; average age 20) in the United States, and Kyoto University (Kytoto; n = 40; 18 Females; average age 20) in Japan. Participants recruited from Kansas received course credit for their participation and participants from Kytoto were given bookstore credits for their participation. All participants had their vision tested using the Freiburg Acuity and Contrast Sensitivity Test (FrACT; Bach, 2007) to ensure they had normal or corrected to normal 20/30 vision and gave their informed consent before participating in the study.

Cultural demographics questionnaire

All participants completed a cultural demographic questionnaire after the experiment. We used the 2010 Census Race and Hispanic Origin Alternative Questionnaire Experiment (AQE; Compton, Bentley, Ennis, Rastogi, 2013) in our study to collect the most detailed information from our participants. This questionnaire collected information about participants' ethnicity, country of origin, and how long they had lived in the country. This enabled us to

2. [2] This definition of the *tyranny of film* hypothesis is a bit different from that used in the past. The original hypothesis was modified in this paper to accommodate the research that has been done since the original study by Loschky et al. (2015), namely Lahnakoski et al. (2014), Hutson et al. (2017), and Huff et al. (2017). The definition in the original study was not as inclusive as this new definition would imply by saying "large differences in cognition that have little or no influence on attention during film viewing," which includes many different factors (i.e., task manipulations, cultural background, context manipulations, attitudes, etc.).

collect detailed information on each participant's cultural and ethnic background, as this was one of our independent variables.

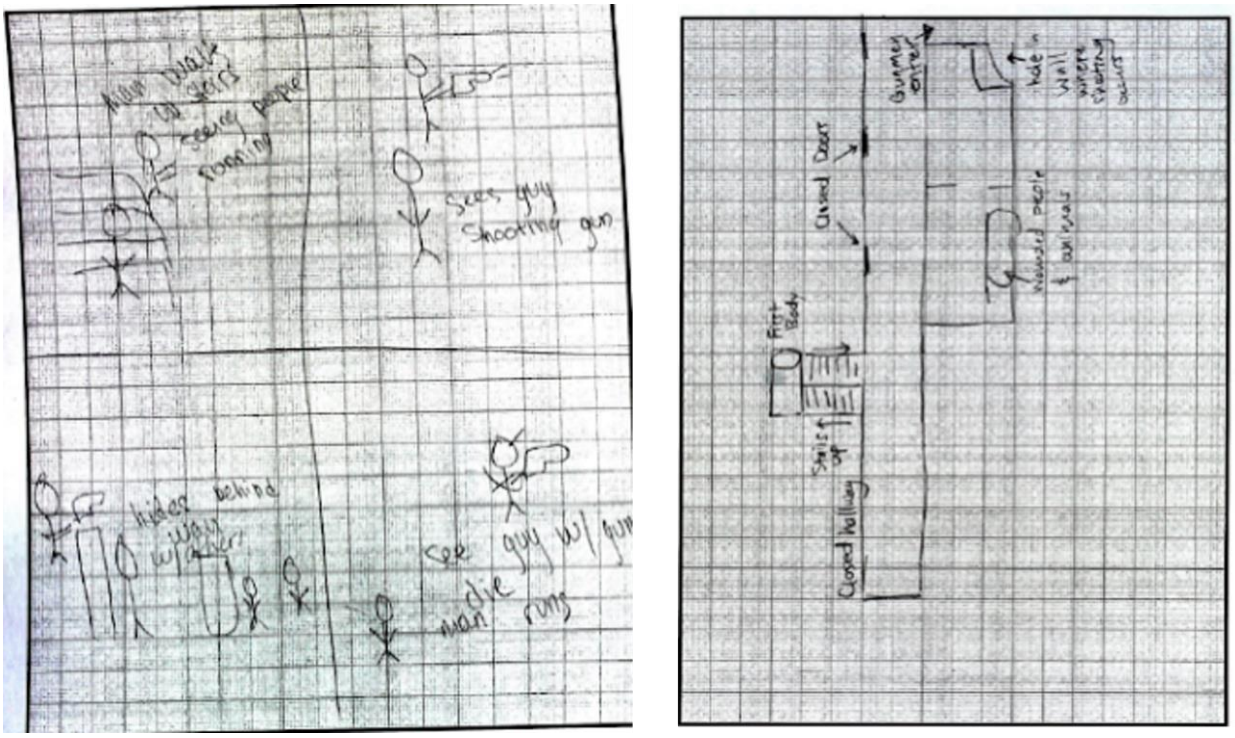
Design

Conditions

An example of the type of responses produced by participants can be seen in Figure 1.2 below. Participants were assigned to one of two between-subject conditions: the map task condition or the comic task condition. The map task was adopted from a study by Hutson et al. (2017) and required participants to draw and label a map depicting as many objects and locations as possible from each film from memory. Participants had five minutes to complete each map, and this was our manipulation of volitional control (i.e., participants must complete the goal of drawing a map from memory). The comic drawing task condition required participants to watch the film for comprehension of the main narrative elements occurring. In this condition, participants drew a 4-panel comic of each of the 8 experimental film clips. The comic condition served to reinforce comprehension of the main narrative elements in the film and kept participants from progressing through the experiment faster than those participants in map task condition, who had to spend time drawing maps. Importantly, the comic task served as a proper control condition in our experiment, which was not something that was present in the Hutson et al. (2017) study. Both tasks were created in a way to be as similar in nature as we could manage. For example, in both tasks, participants drew pictures and wrote words after viewing each clip, had 5 minutes to do so, and were told their task before the beginning of each clip. This ensured that any differences between the two conditions would not be due to small avoidable differences in the tasks itself.

Figure 1.2.

Example Comic and Map Task Figure



Notes. Example responses for the Comic task and Map task from the same film clip (from Children of Men, Cuaron et al., 2006) [left] Example of a comic drawn in the Comic task. The participant drew the required 4-panels and described what was happening in each.

Materials

Computers & eye-tracking

The experiment was programmed in Experiment-Builder (SR Research) and the only differences between Kansas, and Kyoto were in the language of instruction screens (English vs. Japanese). Two senior authors (each a native speaker of English or Japanese, and highly proficient in the other language) checked all translations and made revisions as deemed necessary. All stimuli were presented on a 17'' ViewSonic Graphics Series CRT monitor (Model G90fb) in Kansas and a 19.4'' Iiyama S103MT VisionMaster 503 CRT in Kyoto.

Eye-tracking at both data collection sites was conducted using EyeLink 1000 (SR Research) eye-trackers, with default settings used, including sampling at 1000 Hz (1000 samples per second; SR Research). A chin and forehead rest were used during eye-tracking to maintain a constant viewing distance from the screen (Kansas = 25.2" = 64 cm; Kyoto = 22.44" = 56.99 cm). All participants went through a 9-point calibration and validation process before beginning the procedures of the study. After calibration and before beginning each video, a drift correction was used to ensure participants would maintain focus at the center of the screen and that the calibration was accurate throughout the study. Specifically, as per SR Research guidelines, a maximum error of 1° or better and an average spatial accuracy of 0.5° visual angle were obtained for all calibrations and maintained throughout the study.

Film clips

Eleven different film clips were used in the experiment. The clips were between 1:00 and 3:30 minutes in length and were all unedited, Hollywood style long-shots[3]. The clips were presented at a rate of 30fps and a resolution of 1280 X 1024 pixels at both Kansas (screen size = 14.38'' X 10.75" = 36.53 cm X 27.31 cm) and Kyoto (screen size = 15.5'' X 11.6" = 39.37cm X 29.46 cm). The screen subtended 31.48° X 24.08° of visual angle in Kansas and 36.11° X 28.98° of visual angle in Kyoto. Clips were chosen based on two criteria: 1) there needed to be enough background landmarks, objects, and different locations to ensure that participants in the map task

3. [3] The reason for using only long-shots from these films is that Hollywood-style editing typically uses short shots that, on average, occur every 2-4 seconds. This has been shown to strongly affect eye-movements, because viewers almost invariably move their eyes to the center of the screen immediately after a new shot appears, thus artifactually increasing viewers' Attentional Synchrony (Dorr, et al., 2010; Smith, Levin, & Cutting, 2012; Smith & Mital, 2013).

condition remained engaged with them, and 2) there needed to be enough main narrative elements in the film clip to ensure that participants in the comic task condition remained engaged with them. Of these 11 clips, three were used for practice trials, to ensure that the participants were familiar with and able to perform their respective tasks. The remaining eight clips were used for the experimental portion of the study and were counterbalanced across participants and trials to avoid order or task effects. The source films were "Birdman (from Birdman, Iñárritu et al., 2014)," "Touch of Evil (from Touch of Evil, Welles et al., 1958) ," "James Bond: Spectre (from James Bond: Spectre, Mendes et al., 2015)," "Children of Men (2 clips; from Children of Men, Cuaron et al., 2006)," "Rope (from Rope, Hitchcock et al., 1948)," "The Russian Ark (from The Russian Ark, SoKyotorov et al., 2002)," and "Sacrifice(from Sacrifice, Tarkovsky et al., 1986)." Details on each of the eight experimental clips are given in the Supplemental Materials.

Executive working memory measure

Participants performed the Operation Span (OSPAN) task, which is designed to test the executive-working memory (E-WM) capacity of participants (Bailey, 2012; Conway et al., 2005). The OSPAN task required participants to remember words while they mentally evaluated the truth of math equations. Participants had 4 seconds to determine whether a math equation was true or false. For example, if a participant saw "(10/2) + 3 = 7," they should have responded "incorrect." After participants responded to each math problem, a word was flashed on the screen for 1 second. Participants completed a random number of math equation trials (between 3-7) within a block and were then instructed to recall as many words as they could from within the block (15 blocks total).

Cultural controls

The same experiments were run at both Kansas and Kyoto. All aspects of the study were discussed in detail, to ensure that the two studies were identical to one another between the two laboratories. All the instructions were translated by graduate students at Kyoto who speak both English and Kanji and were checked by professors at Kansas and Kyoto who speak both languages well, to keep the meaning and phrasing as close as possible within both languages.

Procedure

Participants were randomly assigned to 1 of 2 between-subject task conditions (map or comic task) before entering the lab. After giving their informed consent, participants had their vision tested. They then went through a nine-point eye-movement calibration and validation for the EyeLink 1000.

After calibration, participants watched the three practice videos. The first video was used to practice the task for the condition they were randomly assigned (map or comic task). The second practice video was to have participants practice the secondary auditory cognitive load task by itself (more details about the secondary auditory cognitive load task are given in the Supplementary Materials). The third video was to have participants practice the dual-task paradigm (i.e., the map or comic task together with the cognitive load task).

A visual schematic of the main experiment can be seen in Figure 1.1 above. After the practice film clips, participants completed the 8 experimental trials. In those, they watched a video in preparation to draw and write either a map or 4-panel comic from memory afterwards. On half of the trials, either the first or second four film clips, participants completed a second within-subjects manipulation (cognitive load) with 2 levels [cognitive load present or absent] for

the duration of the film clip (see Supplementary Materials for more details about the cognitive load measure used) [1].

After watching all the film clips, participants wrote free recall summaries of the main narrative elements that occurred within each of the film clips. Participants were prompted with two frames from each of the film clips, one from the first second of the film clip and one from the last second of the film clip. This had two purposes: 1) to serve as a retrieval cue for the correct video related information, 2) to minimize the amount of information the participant was given as retrieval cues, to avoid giving away map and comic task-related information. Finally, participants then completed the OSPAN working memory task and the cultural demographic questionnaire. After completing all the above, participants were debriefed and thanked for their participation.

Results

We will first present the results from our Kansas participants, then the results from our Kyoto participants, followed by our cultural comparison analyses[4]. Analyses were conducted using R statistical software (version 3.1.1). We used the lme4 (Bates, Mächler, Bolker, & Walker, 2014) library to run our mixed models, the emmeans (Lenth, 2018) library to plot least squares predicted means from the model fit, the afex (Singmann, Bolker, Westfall, Højsgaard, & Fox, 2015) library to get parameter specific *p* values from the models, and the ggplot2 (Wickham, 2016) library for figure creation.

[4] Data reported in the main body of the paper include only the data where there was no cognitive load present. For detailed results on the full range of data see the Supplementary Materials.

Dependent measures

We chose to analyze the several eye-movement variables, for the following reasons. Eye fixation durations are a standard eye-movement measure of perceptual and conceptual processing time and have been found to differ between East Asian and Western cultures in several studies (Masuda & Nisbett, 2001; Nisbett & Masuda, 2003; Chua, Boland, & Nisbett, 2005). Saccade lengths are a standard measure of the extent of covert attention prior to moving the eyes (Deubel & Schneider, 1996; Hoffman & Subramaniam, 1995; Kowler et al., 1995), and are typically found to inversely co-vary with fixation durations over viewing time in scenes (Antes, 1974; Galpin & Underwood, 2005; Unema et al., 2005; Pannasch et al., 2008). Thus, this would be useful to compare across cultures, given the reported cultural differences in fixation durations. Saccade lengths are also useful in testing the task condition hypothesis. Deviation from screen center is also understood to measure the viewer's breadth of attention for the same reasons as saccade lengths (Loschky et al., 2014; Miura, 1986; Recarte & Nunes, 2003; Reimer et al., 2012). This is important for testing both the Task condition and Cultural dependence hypotheses. The narrative AOI measures (dwell time and fixation count) can both be understood to measure the amount of attention paid to the central narrative elements. Thus, both are important for testing the Task condition hypothesis, which predicts more attention to those narrative elements by participants in the Comic task than those in the Map task (Hutson et al., 2017; Lahnakoski et al., 2014). Likewise, both narrative-AOI measures are important for testing the Culture dependence hypothesis, which predicts more attention to the central narrative-AOIs by those participants from a Western culture than an East Asian culture (Masuda & Nisbett, 2001; Nisbett & Masuda, 2003; Chua, Boland, & Nisbett, 2005; Goh, Tan, & Park, 2009).

Data preparation

Eye-movement data were cleaned to remove outliers. Based on normal ranges for fixation durations during scene perception (for reviews see Fry, 1970; Rayner, 1998) any fixation durations above 3,000 ms (3 seconds) or below 40ms were excluded from the analyses. Saccade lengths that were larger than the display would allow (namely the screen diagonal: 39° in Kansas; 47° in Kyoto), or 0°, were excluded from the analyses. Saccade lengths were then standardized across the two labs for making direct comparisons by dividing the degrees by half of the diagonal (19.96° in Kansas; 23.94° in Kyoto) to create a proportion of maximum possible saccade length that the saccade traveled. The measure of deviation from screen center used the fixation x and y locations relative to the computer screen center and was measured in degrees of visual angle. This measure used the size of the screen diagonal to create a measure in degrees and then to make comparisons across cultures the degree measure was divided by half the diagonal to create a proportion of the maximum possible distance from screen center of each deviation. Narrative areas of interest (narrative-AOIs) were created and used for two separate outcome measures: Dwell Time in narrative-AOIs and number of fixations in the narrative-AOI. These narrative-AOIs were placed on content that was relevant to following the narrative of each film clip, as judged by both the lead author, and confirmed by several research assistants.

Model creation

Generalized-multilevel models (GMLM) were used to analyze all data. We used a Gamma regression to account for the highly positive skew found for continuous eye-movement measures (Lo & Andrews, 2015). For narrative-AOIs fixation count, a GMLM was set to a Poisson function, because it is a discrete outcome measure. Either a log link function or a square root function was used for each model, the lowest AIC and BIC (Schwarz, 1978) values for each

model were used to determine the best model fit for the data. Condition type [map = 1; comic = 0] was entered as the only fixed effect for Experiment 1 and 2 and all models entered participant and video as random effects at their intercept.

Experiment 1: Kansas

Model creation

The guidelines discussed above were used when creating the models in this section. A log link function was used for fixation durations and both narrative-AOIs measures; a square root function was used for Deviation from screen center and saccade amplitudes.

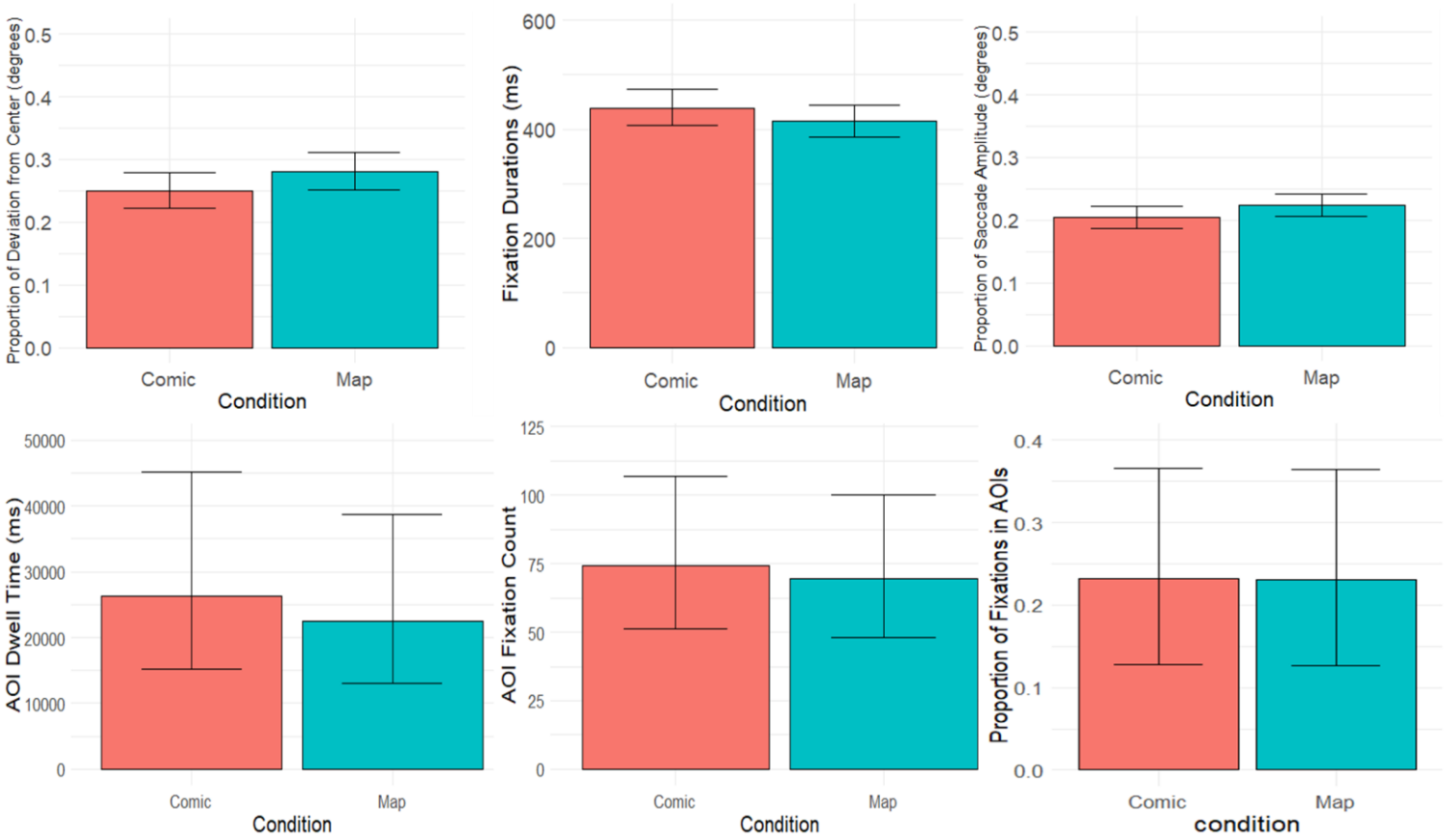
Results

Results from the GMLM are depicted in Figure 1.3 and show several eye-movement measures converging on the same result: support for the *volitional attention hypothesis* in the direction that we predicted with the Map task leading to more exploratory peripheral eye-movements and the Comic task producing more typical eye-movements for film viewing, focusing on the central narrative elements. Condition type was a significant predictor for proportion of maximum deviation from screen center, $\beta = 0.029$, $t = 3.96$, $p < .001$; proportion of maximum saccade amplitude, $\beta = 0.452$, $t = 2.16$, $p = .03$; and narrative-AOIs dwell time, $\beta = 10.176$, $t = -2.58$, $p = .009$. Other measures for which the *volitional attention hypothesis* was refuted, and the *tyranny of film hypothesis* was supported were when predicting fixation durations, $\beta = 6.08$, $t = -1.19$, $p = .236$, proportion of fixations made to narrative-AOIs, $\beta = 0.481$, $t = -0.154$, $p = .878$, and narrative-AOIs fixation count, $\beta = 4.31$, $Z = -1.17$, $p = .243$, which showed no difference between the two conditions (though there was a non-significant trend for longer fixations in the Comic task than the Map task). This pattern of results suggests that participants in the Map condition did not disengage from the narrative completely, but rather

they still attended to the narrative (e.g., no difference in the proportion of fixations to the narrative-AOIs) but did not look for as long as those in the comic condition (e.g., shorter dwell times in the narrative-AOIs in the map task). Means and standard errors for each model can be found in Table 1.1 below.

Figure 1.3.

Kansas Eye-Movement Figures



Notes. All values represent the raw data and error bars represent 95% confidence intervals. Results of the GMLM where the outcome variable was [top left] the proportion of maximum deviation from screen center (degrees). [top middle] fixation durations (ms). [top right] proportion of maximum saccade amplitude (degrees). [bottom left] the dwell time within the narrative-AOIs (ms). [bottom middle] the number of fixations made to the narrative-AOIs, [bottom right] proportion of fixations made to narrative-AOIs.

Table 1.1.

Task Condition Analyses (Kansas)

Values depicting the means, standard errors, and significant values for each of the outcome measures

Predictor	*M*	*SE*	*p*
Proportion of Deviation*			
Comic Task	0.50	0.01	<.001
Map Task	0.53	0.01	
Proportion of Saccade lengths			
lengths*	0.45	0.01	.03
Comic Task	0.47	0.01	
Map Task			
Fixation Durations			
Comic Task	6.08	0.4	.236
Map Task	6.03	0.4	
Narrative-AOI Dwell Time*			
Comic Task	10.2	0.28	.009
Map Task	10	0.28	
Narrative-AOI Fixation Counts			
Comic Task	4.31	0.19	.243
Map Task	4.24	0.19	
Proportion Narrative-AOI Fixations			
Comic Task	0.481	0.063	.878
Map Task	0.480	0.063	

*Note: * indicates a significant difference at the p<.05 level*

Discussion

The eye-movement results support the *volitional attention hypothesis*. Specifically, we hypothesized that participants in the Map task would pay less attention to the central narrative AOIs. Specifically, the map task was argued to be at odds with comprehending the narrative, since it is focused on remembering landmarks and their relative spatial locations and not the narrative events. Likewise, we hypothesized that in the comic task condition (i.e., a control

condition for the Map task), which was focused on comprehending the narrative, participants would attend more to the main characters in the film narrative, and less time attending to background information. The results further show participants doing the Map task made longer saccade lengths, and had greater deviation from screen center, while participants doing the Comic task had shorter saccade lengths and less deviation from screen center. These results merge nicely with the narrative-AOIs viewing time results, which were shorter in the map task. However, the proportion of fixations in the narrative-AOI results also suggest that the map task participants did not completely disengage from the narrative content (i.e., there were similar proportions of fixations to the narrative-AOIs in both tasks). This suggests that the map task participants were not able to completely tune out the film narrative even when given a task at odds with comprehension, which may suggest some tyranny of film effects remained. Generally, these results both replicate the results of Hutson, et al. (2017), and extend them from one to eight film clips (with the sole exception being for fixation durations). Thus, the finding that a volitional attention task at odds with comprehending the film narrative can break the Tyranny of film, is generalizable across a range of different film clips.

While these results were consistent with our predictions, we wondered if we would replicate them in both a different laboratory, and more importantly, a different culture. As described above, some research suggests that East Asian and Western cultures have different attentional selection patterns, which specifically differ between attention to central focal elements versus scene backgrounds (Chua, Boland, & Nisbett, 2005; Goh, Tan, & Park, 2009; Masuda, Ishii, & Kimura, 2016). Given that our task manipulations also concern viewers' attention to central focal elements versus backgrounds, this becomes an even more interesting question. Finally, given that any cultural difference in attentional selection would necessarily be

due to experience (Karden et al., 2017), and thus a mandatory top-down influence on attention, it is possible that it would be wiped out by the Tyranny of Film, as we have previously found for the mandatory top-down effect of the film viewer's event model. Thus, Experiment 2 was run to test these hypotheses. The only change made from Experiment 1 to Experiment 2 was the culture where participants' data was collected.

Experiment 2: Kyoto

Model creation

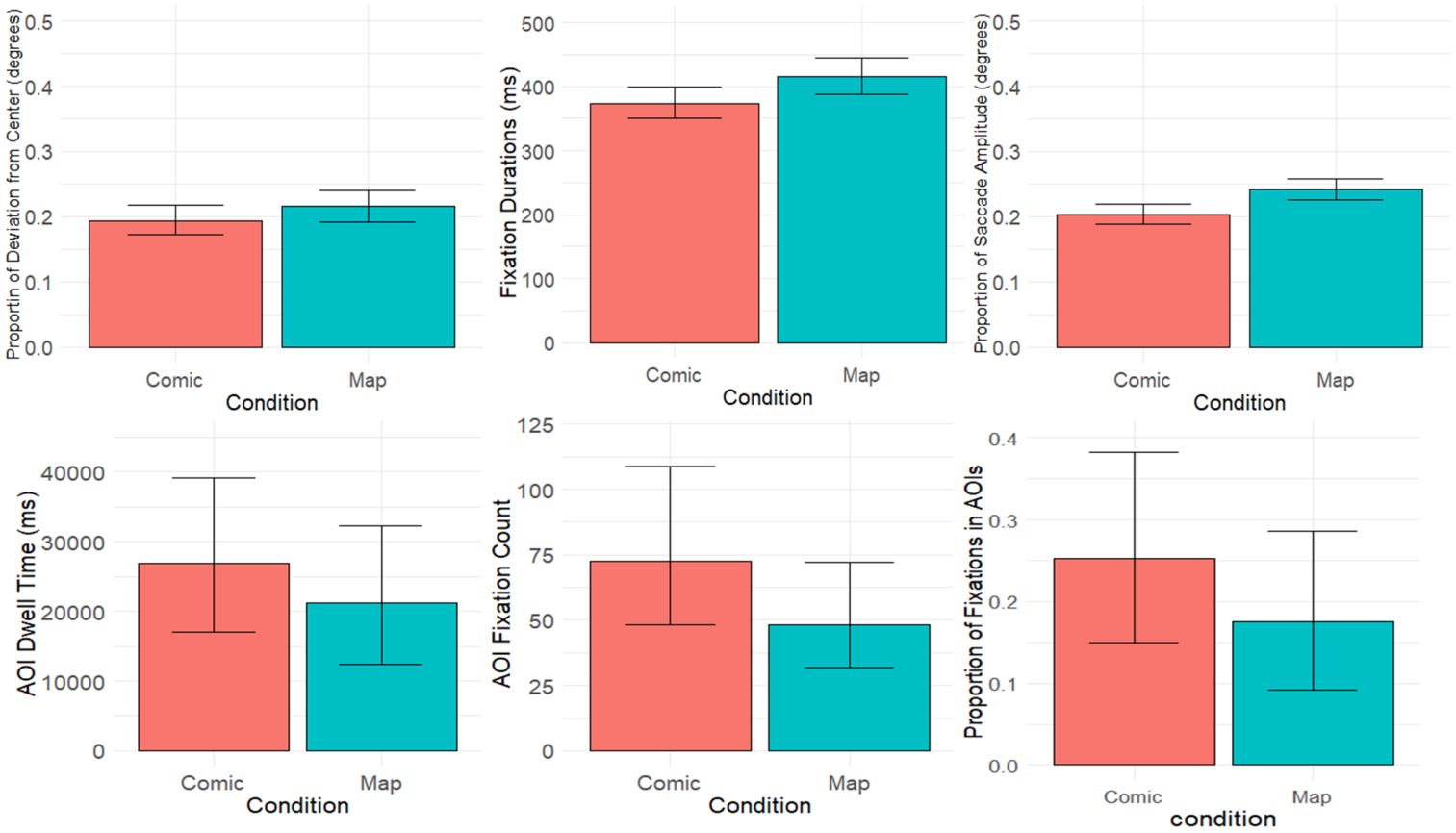
The guidelines discussed above were used when creating the models in this section. A square root link function was used for all eye-movement measures for this set of data analyses.

Results

The results for Kyoto are visually depicted in Figure 1.4 below and are almost entirely in line with the Kansas results, in that they are consistent with the *volitional attention hypothesis* across our range of eye-movement measures: proportion of maximum deviation from screen center, $\beta = 0.44$, $t = 3.11$, $p = .002$; proportion of maximum saccade amplitudes, $\beta = 0.45$, $t = 4.88$, $p < .001$; and narrative-AOIs Dwell time, $\beta = -18.54$, $t = -2.43$, $p = .015$. The only differences in the task effects between the two cultures were that participants in Kyoto showed significant effects of condition for *all* eye-movement measures, including fixation durations, $\beta = 19.34$, $t = 2.72$, $p = .007$, proportion of fixations made to the narrative-AOIs, $\beta = 0.42$, $t = -9.46$, $p < .001$, and narrative-AOIs fixation counts, $\beta = -0.409$, $Z = -4.76$, $p < .001$. Means and standard errors can be found in Table 1.2 below for each predictor.

Figure 1.4.

Kyoto Eye-Movement Figures



Notes. All values represent the raw data and error bars represent 95% confidence intervals. Results of the GMLM where the outcome variable was [top left] the proportion of maximum deviation from screen center (degrees). [top middle] fixation durations (ms). [top right] proportion of maximum saccade amplitude (degrees). [bottom left] the dwell time within the narrative-AOIs (ms). [bottom middle] the number of fixations made to the narrative-AOIs, [bottom right] proportion of fixations made to narrative-AOIs.

Table 1.2.

Task Condition Analyses (Kyoto)

Values depicting the means and standard error for each of the predictor variables tested.

Predictor	*M*	*SE*	*p*
Proportion of Deviation*			
Comic Task	0.44	0.01	.002
Map Task	0.46	0.01	
Proportion of Saccade lengths lengths*			
Comic Task	0.45	0.01	<.001
Map Task	0.49	0.01	
Fixation Durations*			
Comic Task	19.24	0.3	.007
Map Task	20.40	0.4	
Narrative-AOIs Dwell Time*			
Comic Task	9.87	0.16	.015
Map Task	9.75	0.16	
Narrative-AOIs Fixation Counts*			
Comic Task	4.28	0.21	<.001
Map Task	3.87	0.21	
Proportion Narrative-AOI Fixations*			
Comic Task	0.502	0.059	<.001
Map Task	0.419	0.059	*

Note: * signifies a significant difference at the p<.05 level

Discussion

Importantly, the results of Experiment 2 replicated the Volitional Attention Hypothesis results found in Experiment 1, in a different laboratory, and more importantly, a different culture. We found that the task condition influenced attentional selection during film viewing, breaking the Tyranny of Film. Specifically, the Map task increased fixation durations, proportion of saccade lengths, and proportion of deviation from screen center, relative to the Comic task. These results pair nicely with the narrative-AOIs results that show in the Map task participants

made less fixations and spent less time dwelling in the narrative-AOIs. What is less clear, however, is whether the results are significantly different between the two cultures or if both cultures were showing similar impacts of the different task manipulations.

Experiment 2a: Location Comparison

All the descriptive statistics for the raw eye-movement data are reported in Table 1.3 by location. Overall, Kansas participants had longer fixation durations, more deviation from screen center, longer narrative-AOI dwell times, and more fixations to narrative-AOIs. These descriptive results are all broadly consistent with the *Culture dependence* hypothesis.

Table 1.3.

Descriptive Statistics: Location Comparison

Descriptive Statistics for the six eye-movement measures broken up by Location of participants

Measure	*M*	*SD*
Fixation Duration (ms)		
Kansas	418.73	331.32
Kyoto	386.39	298.09
Proportion of Saccade length (degrees)		
Kansas	0.219	0.19
Kyoto	0.215	0.19
Proportion of Deviation (degrees)		
Kansas	0.278	0.18
Kyoto	0.205	0.13
Narrative-AOI Dwell Time (ms)		
Kansas	26,571	21,715
Kyoto	20,964	17,390
Narrative-AOI Fixation Count		
Kansas	76	64
Kyoto	67	56
Fixation Count		
Kansas	281	114
Kyoto	292	116
Proportion of Narrative-AOI Fixations		
Kansas	.202	0.122
Kyoto	.19	.13

Model creation

The fixed effects for the following models were condition type [map = 1; comic = 0], location [Kyoto = 1; Kansas = 0], and the Location X Condition Type interaction. We entered participant and video as random effects at their intercept. A log link function was used for

fixation durations, saccade amplitudes, narrative-AOIs fixation count and narrative-AOIs dwell time, while a square root link function was used for deviation from screen center. These link functions were determined based on which produced the better AIC and BIC values for each model.

Results

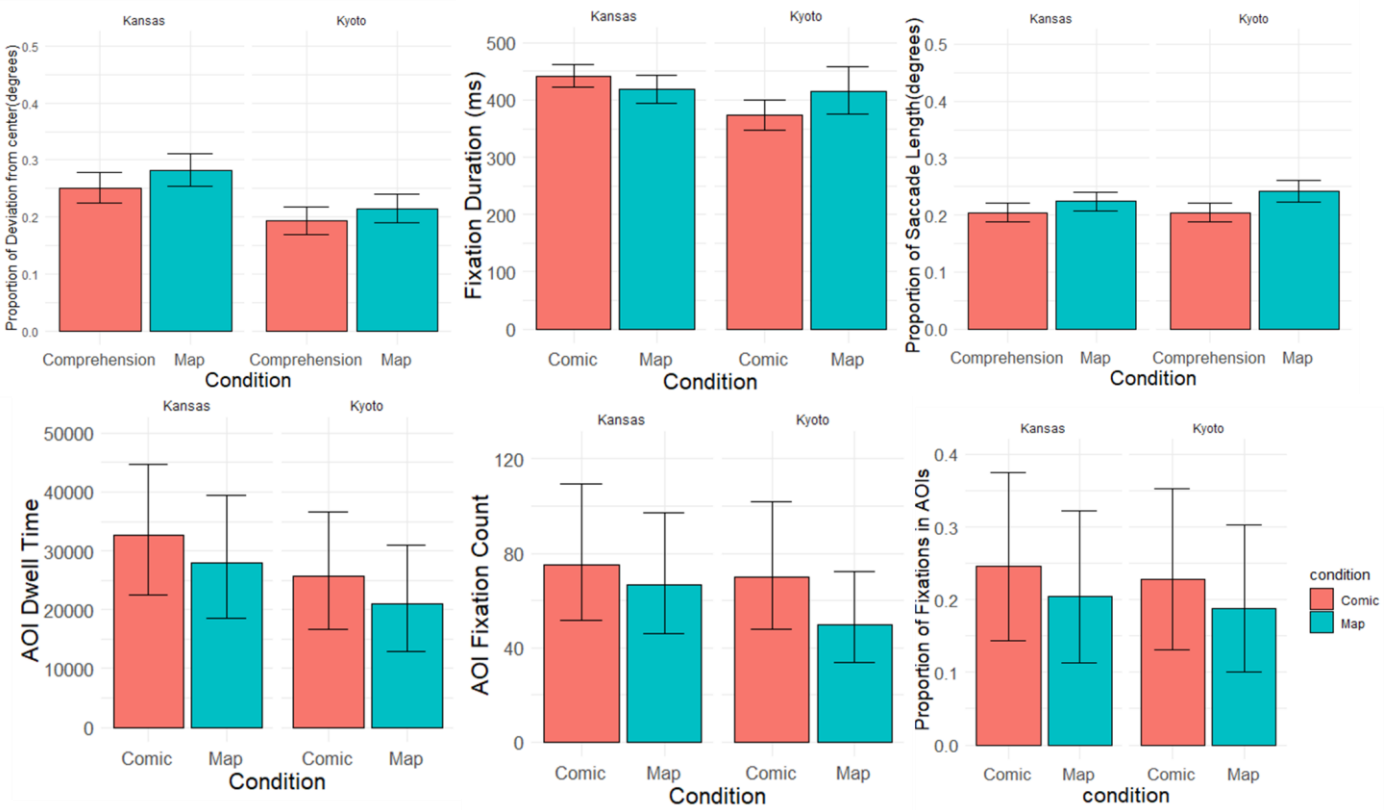
Here we report results for the Location and Location X Condition interaction only. Since the Condition effects were discussed separately above, we will not repeat them here (though they can be found in the Supplementary Materials). Results from a GMLM are depicted in Figure 1.5 and show support for the *cultural dependence hypothesis* for all our measures (i.e., proportion of deviation from screen center, $\beta = 0.501$, $t = -7.89$, $p < .001$; fixation duration, $\beta = 19.33$, $t = -5.596$, $p < .001$; narrative-AOIs dwell time, $\beta = 9.64$, $t = -39.701$, $p < .001$; narrative-AOIs fixation count, $\beta = 3.92$, $Z = -80.51$, $p < .001$, and proportion of fixations made to narrative-AOIs, $\beta = 0.472$, $t = -2.379$, $p = 0.17$) expect one (proportion of saccade amplitudes $\beta = -1.596$, $t = 0.021$, $p = .983$). The means and standard errors for each of these predictors can be seen in the Table 1.4 below, though generally Kyoto had more exploratory eye-movements while Kansas had more focal eye-movements.

Additionally, results from the GMLM show some support for the Location X Condition interaction: fixation durations, $\beta = 22.649$, $t = 6.027$, $p < .001$; narrative-AOIs Dwell Time, $\beta = 9.87$, $t = -2.691$, $p = .007$; and narrative-AOIs fixation counts, $\beta = 3.81$, $Z = -110.99$, $p < .001$; with no significant interaction effects for deviation from screen center or saccade lengths. This suggests that the eye-movements in each cultural location are being differentially impacted by the two task conditions in this study.

Specifically, there was a cross-over interaction for fixation durations, such that Kyoto participants had longer fixations in the map condition and Kansas participants had longer fixations in the Comic condition. Additionally, there were magnitude interactions for the analyses related to narrative-AOI (i.e., narrative-AOI Dwell Time, narrative-AOI fixation count, and proportion of fixations to AOIs), where these results largely were in-line with the other cultural differences already discussed. Specifically, Kyoto had eye-movement patterns consistent with relatively higher degrees of exploration overall, but especially in the map condition. There were no a priori hypotheses regarding a Location X Condition interaction, though these results seem to be mostly in line with what we would have expected based on our other predictions. Specifically, since the map task condition requires participants to use more exploratory viewing patterns to complete the task, it would make sense that Kyoto participants would be during even more so under these circumstances. The one result that does not fall in line with this general idea is the proportion of deviation from screen center. One possibility is that this result could be a bi-product of the film clips that were used (e.g., all Western culture style film clips) impacting the viewing patterns of Kyoto participants to not deviate as much from the center of the screen (Ueda & Komiya, 2012), though Kyoto participants still showed signs of being more exploratory overall. These results are more clearly depicted in the form of a set of averaged heat maps in the Supplementary materials of the paper.

Figure 1.5.

Cultural Comparison Eye-Movement Figures



Notes. All values represent the raw data and error bars represent 95% confidence intervals. Results of the GMLM where the outcome variable was [top left] the proportion of maximum deviation from screen center (degrees). [top middle] fixation durations (ms). [top right] proportion of maximum saccade amplitude (degrees). [bottom left] the dwell time within the narrative-AOIs (ms). [bottom middle] the number of fixations made to the narrative-AOIs, [bottom right] proportion of fixations made to narrative-AOIs.

Table 1.4.

Task Condition Analyses (Cultural Comparison)

Values depicting the means and standard error for the cultural predictor model outputs.

Predictor	*M*	*SE*	*p*
Proportion of Deviation*			
Kansas	0.52	0.01	<.001
Kyoto	0.45	0.01	
Proportion of Saccade			
Kansas	0.47	0.01	.983
Kyoto	0.46	0.01	
Fixation Durations*			
Kansas	3.31	0.04	<.001
Kyoto	3.07	0.04	
Narrative-AOIs Dwell Time*			
Kansas	9.82	0.011	<.001
Kyoto	9.55	0.011	
Narrative-AOIs Fixation Counts*			
Kansas	3.95	0.10	<.001
Kyoto	3.77	0.10	
Proportion of Narrative-AOI Fixations*			.017
Kansas	0.487	0.06	
Kyoto	0.443	0.06	

Note: * signifies a significant difference at the p<.05 level

Discussion

The results of experiment 2a provided strong support for the *Cultural Dependence hypothesis*, and rejection of the *Tyranny of Film hypothesis*. Namely, when viewing film clips there were cultural differences in eye-movements between our Kansas and Kyoto samples. Specifically, first, for fixation durations, Kyoto participants were shorter, while Kansas participants were longer. This replicates the most consistent difference in eye-movements between East Asian and Western cultures (Chua, Boland, & Nisbett, 2005; Goh, Tan, & Park,

2009; Masuda, Ishii & Kimura, 2016). Conversely for saccade lengths, we did not find a cultural difference between Kansas and Kyoto participants. Notably, this cultural difference has only been indirectly shown previously in a single study comparing East Asian and Western cultures (Goh, Tan, & Park, 2009). Concerning deviation from screen center, we found that Kansas participants had more deviation, while Kyoto participants had less. Only one previous study has measured this variable (Goh, Tan, & Park, 2009), and they found the opposite result. It is not immediately obvious why our results differed from theirs. However, concerning fixation counts, proportion of fixations, and dwell times in the narrative-AOIs, the Kyoto participants had fewer fixations, a lower proportion of fixations, and less time spent in Narrative-AOIs, than Kansas participants. This is consistent with several prior studies that have either shown that East Asians made relatively fewer fixations on central items than Westerners (Chua, Boland, & Nisbett, 2005; Masuda, Ishii, & Kimura, 2016), or that East Asians less time spent looking at central items than Westerners (Masuda, Ishii, & Kimura, 2016; Rayner, Li, Williams, Cave & Well, 2007; Senzaki, Masuda & Ishii, 2014).

These results are important, not only for showing clear cultural effects on eye-movements, but also because it has been shown in the context of viewing dynamic scenes (i.e., film). This shows it is possible for bottom-up features of a film to be overridden by the mandatory top-down influences of culture. Thus, the current study shows two very different sorts of top-down attentional influences, volitional tasks at odds with comprehending the film narrative, and mandatory cultural influences, that can break the tyranny of film.

General Discussion

The aim of this study was threefold. Firstly, we wanted to replicate and extend findings that the tyranny of film can be overridden while viewing film with a task manipulation that is at

odds with comprehending the film narrative (Hutson, et al., 2017). Secondly, we wanted to investigate the highly contentious question of cross-cultural differences in eye-movements which has produced widely disparate results in prior studies (Masuda & Nisbett, 2001; Nisbett & Masuda, 2003; Chua, Boland, & Nisbett, 2005; Goh, Tan, & Park, 2009; Rayner et al., 2007; Rayner, Castelhano, & Yang, 2009; Ueda & Komiya, 2012; Masuda, Ishii, & Kimura, 2016). Thirdly, we wanted to test if whether a volitional attention task at odds with comprehending a film is cognitively demanding. This last question is not discussed here, but instead in the supplementary materials, and is much further explored in a separate study (Simonson et al., in preparation).

Executive function discussion

We replicated the results of Hutson et al. (2017) with new participants in two different laboratories and extended it from one to eight different film clips. These are both crucial points in making this study's findings more generalizable. Additionally, by including the Comic task as a control condition, we also showed that a task with a goal (e.g., the comic task) is not sufficient to break the tyranny of film, but rather the task must also be at odds with narrative comprehension (e.g., the map task). We induced volitional attentional selection using the map task paradigm and found significant condition effects across experiments. Specifically, those participants who were in the comic task condition generally had comparatively shorter saccade lengths, longer fixations, decreased deviation from screen center, more dwell time in the narrative-AOIs, and more fixations to the narrative-AOIs on average when compared to the participants in the map task condition. This replicates the results of previous studies, showing a task at odds with comprehension of a film, shows significant difference in eye-movements

(Lahnakoski et al., 2014; Hutson et al., 2017), which lessens the tyranny of film that has previously been shown when viewing film clips (Loschky et al., 2015).

Culture discussion

Regarding the cultural motivations, there was support for the Culture Dependence hypothesis. This was the hypothesis that we would find differences in attentional selection between an East-Asian and a Western culture while watching film. Testing this hypothesis was important for two reasons: 1) research has been very mixed in the past regarding if there are cultural differences in attentional selection and 2) it would further test whether bottom-up features of the film can be overridden by top-down factors (i.e., cultural influences), namely evidence for overriding the tyranny of film. Our results showed that Kyoto participants attended relatively more to the background, and less on central narrative elements (i.e., main characters or objects, usually near the center of the screen), while the opposite was true for Kansas participants who attended relatively more to the central narrative elements of the film clips.

Interestingly, there were two results in our cultural comparison that require a more nuanced discussion. Specifically, we found that Kyoto participants had a lower proportion of deviation from screen center than Kansas participants, and we found no difference in proportional saccade lengths across the two cultures. These results need to be squared with the differences we found between the two cultures in their number and proportion of fixations in the central narrative-AOIs, and time spent looking at them. As noted above, for the latter three measures, Kyoto participants attended less to these central narrative elements than Kansas participants. Together, the two sets of results suggest that although Kyoto participants deviated less from the screen center than the Kansas participants, they still spent less time focusing on the central characters and objects, generally found near the center of the screen. This, in turn,

suggests that Kyoto participants instead explored background elements that were nevertheless closer to the center of the screen. This explanation would be in line with previous research that found East Asian participants were more exploratory than Western participants, perhaps to conceptually relate more elements in a scene to each other (Nisbett & Masuda, 2003).

An interesting question is why we found such a wide range of cultural differences between East Asians' and Westerner's eye movements in our study, while past research on this question has produced highly mixed results. First, it may be a bi-product of having greater statistical power in our experiments. Specifically, we had 86 participants, who watched 8 film clips each, which produced over 200 thousand data points. Due to the sheer amount of data, our study may simply have had greater sensitivity to the differences that existed but could not be reliably found in the seven prior studies comparing the eye movements of East-Asian and Westerners. Six of those seven studies measured fixation durations, and three found differences in the same direction that we found (East-Asian < Western; Chua, Boland, & Nisbett, 2005, N = 43; Goh, Tan, & Park, 2009, N = 30; Masuda, Ishii, & Kimura, 2016, N = 101), one study found a marginal difference in the same direction (Rayner et al., 2007, N = 47), and two studies found null effects (Evans et al., 2009, N = 43; Rayner et al., 2009, N = 24). Note that the number of participants in the above studies that showed a cultural difference in fixation durations ranged from 30-101, while those that found null effects ranged from 24-47. Our study, which had 86 participants, likely had greater power than all but one of the prior studies (Masuda, Ishii, & Kimura, 2016), which alone could explain our greater sensitivity in finding the range of cultural differences across different eye movement metrics.

A different explanation for our finding stronger evidence of East-Asian versus Western differences in eye movements is in terms of where we collected our data from our Japanese and

American participants. In our study, we collected data from participants within their own cultures, which not all previous studies did. Rather, most used convenience samples of culturally diverse students from a single institution in the US or Canada (Chua, Boland, & Nisbett, 2005; Evans et al., 2009; Rayner, 2007; Rayner, Castelhano, & Yang, 2009). Going back to our above comparison of East-Asians and Westerners' fixation durations of the three previous studies that found significant culture effects, two collected data from participants in their home cultures (Goh, Tan & Park, 2009: in Singapore vs. the US; Masuda, Ishii & Kimura, 2016: in Japan vs. Canada) and one collected data from participants living in the US (Chua, Boland, & Nisbett, 2005). The study that found a marginal effect collected data from participants living in the US (Rayner et al., 2007). The two studies that found null effects both collected data from participants living in the US (Evans et al., 2009; Rayner et al., 2009). This suggests that participants not exposed to the same cultural location for an extended period may not be ideal for studying cultural differences in attention.

This is suggested by a study by Ueda and Komiya (2012), which showed an effect of cultural scene priming on Japanese participants' visual attention. In a within-subjects design, participants looked at culturally neutral images of single or multiple objects (e.g., a dog, vs. fruits in a bowl) both before and after being primed by looking at a block of >200 cultural photos of urban and suburban street scenes from either Japan, or the US). While viewing the priming scenes, participants fixated more broadly when looking at the Japanese scenes than the US scenes. Furthermore, after priming by Japanese street scenes, when participants first looked at a single neutral object image, their fixations were more broadly distributed than after priming by US street scenes. This suggests that breadth of attentional selection can be influenced by mere exposure to scenes from different cultural environments. Thus, in studies where participants are

not immersed in their own culture, there may be changes in their cultural patterns of attentional selection. This highlights the importance of sampling in a participants' home environment.

Since the video clips that we used were all from Western culture films, it is possible that this had some impact on the results between our two cultures. Specifically, as discussed above, participants can be biased to view a scene in a particular way based on priming by the culture that scenes were taken from (Ueda & Komiya, 2012). Thus, it would be an important next step to use film clips taken from the Japanese or other East-Asian cultures and rerun the experiment to compare the results. This would allow for the comparison between stimuli to identify how the cultural location of the stimuli impacts attentional selection during film viewing.

Interestingly, the cultural differences results that we are seeing during film viewing could be the result of a culturally-specific form of the tyranny of film. While we have thought of the tyranny of film as a factor inhibiting viewers from exploring film scenes in different ways from other viewers, it is possible that through some culturally-based cognitive and perceptual processes, individual differences were inhibited and thus made viewers' eye-movements more similar within each culture. This could then lead to the differences between cultures that we found. Thus, identifying if participants are more similar to other participants within the same culture, but more different across cultures, would help identify if the tyranny of film is the same for everyone, or is fundamentally different for each culture (e.g., members of Western cultures are more likely to naturally view the central narrative elements, thus that is what their tyranny of film looks like; members of East-Asian cultures are more likely to view the entire scene, thus their tyranny of film would look fundamentally different from Western participants). Under this model, the key question would be to determine if the tyranny of film looks the same or different for each culture. Importantly, however, if one argues for a culturally-specific form of tyranny of

film, then it cannot be strictly based on the film stimulus, given that film clips were held constant.

Further, if the tyranny of film is based solely on the film stimulus, and involves reducing individual differences between viewer's eye movements, then one could argue that using film as our medium in this study could have lessened individual differences within each culture. If so, it could help explain why we have found such strong effects of cultural differences on eye movements when other studies have not (Evans et al., 2009; Goh, Tan, & Park, 2009; Rayner et al., 2009; Rayner et al., 2007; Masuda, Ishii, & Kimura, 2016). Specifically, some other studies have shown no (or minimal) cultural differences in attentional selection when using static images as their stimuli, which may have allowed a stronger role for individual top-down influences on their eye movements, weakening any evidence of cultural differences. Specifically, motion seems to be the single strongest visual stimulus factor in directing viewers' gaze (Mital et al., 2011; Carmi & Itti, 2006), thus static images should allow greater individual differences in where viewers look. Thus, our observed cultural differences in attentional selection may have become more apparent when viewing dynamic video images, due to film weakening individual differences in eye movements within each culture. Nevertheless, this argument involves a key complication. Namely, since the video clip stimuli were held constant across cultures, and we found differences in eye movements across those cultures, this explanation also requires an interaction between stimulus-driven effects and culturally-specific experience-driven effects. Thus, as above, if one argues for a culturally-specific form of tyranny of film, it requires that the tyranny of film not be strictly based on the film stimulus.

Future directions

The current results point to the importance of further investigating differences in attentional between cultures for several reasons. Firstly, it is possible (as stated above) that the tyranny of film was not really broken by culturally-specific attentional selection patterns, but rather our understanding the tyranny of film may need to change. If so, then there may not be a single universal tyranny of film, based solely on the film stimulus. Rather, there could be different forms of the tyranny of film for different cultures, involving systematic interactions between the film stimulus and shared cultural experience that is unconsciously held and highly overlearned. If so, then it would suggest that film viewers' eye-movements are different between cultures, but not within cultures. This is a direction that future research should aim to tease apart further.

Secondly, more research should be done looking into different types of film clip stimuli. In this set of studies, the stimuli that were used were clips taken from Western cultures. While the cultural differences in attention that we show here are reliable, there could be some effect of the stimuli on the overall pattern of our participants' eye-movements. Specifically, it is possible that when East Asian and Western viewers are presented with East Asian film clips, we would find further differences in attentional selection. Namely, East Asian participants' eye movements could show even more within-culture similarity based on greater familiarity. We have assumed that all participants were familiar with the type of filmmaking in Western films, due to their large availability (e.g., through streaming and increased media access), but this may not be the case. At the very least, future research on cultural differences in attention should measure familiarity with the type of stimulus that is being used to better take that into consideration.

Conclusion

In conclusion, we found evidence that the tyranny of film can be reduced during volitional and mandatory attentional selection processes. Specifically, when viewers did a task at odds with comprehension, which required using volitional top-down attention, they produced a different pattern of eye-movements compared to viewers who were just watching the film clip for comprehension. Additionally, we found that the cultural background of the viewers, which evoked mandatory top-down attention, also produced a different patterns of eye movements. This further reduced the tyranny of film and showed stronger evidence of differences in attention between East-Asians and Westerners than in past research that had used static images. These two separate modes of top-down attention, volitional versus mandatory (Baluch & Itti, 2011), have not been separately discussed in most of the prior literature on attentional selection, which could be why differences in eye-movement patterns have not been found as often when using dynamic images (i.e., video stimuli; Lahnakoski et al., 2014; Loschky, et al., 2015; Hutson, et al., 2017). This suggests that further research should differentiate both types of top-down attention, since they each influence attentional selection in distinct ways.

References

Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology, 103*(1), 62-70.

Bach, M. (2007). The Freiburg visual acuity test- Variability unchanged by post-hoc re-analysis. *Graefe's Archive for Clinical and Experimental Ophthalmology, 245(*7), 965-971.

Bailey, H. (2012). Computer-paced versus experimenter-paced working memory span tasks: Are they equally reliable and valid? *Learning and Individual Differences, 22*, 875-81.

Baluch, F. & Itti, L. (2011). Mechanisms of top-down attention. *Trends in Neuroscience, 34*(4), 210-224.

Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Carmi, R., & Itti, L. (2006). Visual cause versus correlates of attentional selection in dynamic scenes. *Vision Research, 46*, 4333-4345.

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye-movements during scene perception. PNAS, *102*(35), 12629-12633.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O, & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769-786.

Compton, E., Betley, M., Ennis, S., & Rastogi, S. (2013) 2010 Census Race and Hispanic Origin Alternative Questionnaire Experiment (No. 211).

DeAngelus, M. & Pelz, J. B. (2009). Top-down control of eye movement: Yanbu's revisited. *Visual Cognition, 17*(6-7), 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).

Evans, K., Rotello, C. M., Li, X., & Rayner, K. (2009). Scene perception and memory revealed by eye movements and receiver-operating characteristic analyses: Does a cultural difference truly exist?, *The Quarterly Journal of Experimental Psychology*, *62*:2, 276-285.

Galpin, A. J., & Underwood, G. (2005). Eye movements during search and detection in comparative visual search. *Perception & Psychophysics, 67*(8), 1313-1331.

Goh, J. O., Tan, J. C., & Park, D. C. (2009). Culture Modulates Eye-Movements to Visual Novelty. *PLoS ONE, 4* (12), e8238.

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

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: Elsevier Science Ltd.

Hutson, J. P., Smith, T. J., Magliano, J. P., & Loschky, L. C. (2017). What is the role of the film viewer? The effects of narrative comprehension and viewing task on gaze control in film. *Cognitive Research: Principles and Implications, 2*(24).

Karden, O., Shneidman, L., Krogh-Jespersen, S., Gaskins, S., Lerman, M. G., & Woodward, A. (2017). Cultural and Developmental Influences on Overt Visual Attention to Videos. *Scientific Reports,* 7: 11264, 1-16.

Lahnakoski, J. M., Glerean, E. J., Jaaskelainen, I. P., Hyona, J., Hari, R., Sams, M., … Nummenmaa, L. (2014). Synchronous brain activity across individuals underlies shared psychological perspectives. *NeuroImage,* 100, 316-324.

Lenth, Russel. (2018). Emmeans: Estimated Marginal Means. aka Least-Squares Means, R package version 1.3.0. https://CRAN.R-project.org/package=emmeans

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyze reaction time data. *Frontiers in Psychology*, 6, 1171.

Loschky, L.C., Ringer, R., Johnson, A., Larson, A.M., Neider, M., & Kramer, A. (2014). Blur detection is unaffected by cognitive load. *Visual Cognition, 22*(3/4), 522-547.

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.

Masuda T., Ishii K., & Kimura J. (2016). When Does the Culturally Dominant Mode of Attention Appear or Disappear? Comparing Patterns of Eye Movement During the Visual Flicker Task Between European Canadians and Japanese. *Journal of Cross-Cultural Psychology, 47*(7), 997-1014.

Masuda, T., & Nisbett, R. E. (2001). Attending Holistically Versus Analytically: Comparing the Context Sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology, 81*(5), 922-934.

Mitchell, J. P., Macrae, N., & Gilchrest, I. D. (2002). Working Memory and the Suppression of Reflexive Saccade lengths. *Journal of Cognitive Neuroscience, 14*(1), 95-103.

Mital, P.K., Smith, T.J., Hill, R.L., & Henderson, J. M. (2011). Clustering of Gaze During Dynamic Scene Viewing is Predicted by Motion, *Cognitive Computation, 3*5-24.

Miura, T. (1986). Coping with situational demands: A study of eye movements and peripheral vision performance. *Vision in vehicles*.

Nisbett, R. E. & Masuda, T. (2003). Culture and point of view. *PNAS, 100*(19), 11163-11170.

Pannasch, S., Hmert, 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.

Rayner K., Castelhano M.S., & Yang J. (2009). Eye Movements When Looking at Unusual/Weird Scenes: Are There Cultural Differences?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(1), 254-259.

Rayner K., Li X., Williams C.C., Cave K.R., & Well A.D. (2007). Eye movements during information processing tasks: Individual differences and cultural effects. *Vision Research, 47,* 2714-2726.

Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology-Applied, 9*(2), 119–137.

Reimer, B. (2009). Impact of cognitive task complexity on drivers' visual tunneling. *Transportation Research Record: Journal of the Transportation Research Board*, 2138, 13–19.

Reimer, B., Mehler, B., Wang, Y., & Coughlin, J. F. (2012). A field study on the impact of variations in short-term memory demands on drivers' visual attention and driving performance across three age groups. *Human factors*, *54*(3), 454-468.

Schwarz, G. (1978). Estimating the Dimension of a Model, *Annals of Statistics, 6*, 461-464.

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). afex: Analysis of Factorial Experiments. R package version 0.16-1. https://CRAN.R-project.org/package=afex

Simonson, T. L., et al. (2020). Executive Control of Eye-Movements in Film: Understanding how the Tyranny of Film is attenuated by cognitively demanding tasks. [Manuscript in preparation]. Department of Psychological Sciences, Kansas State University.

Smith, T. Henderson, J. (2008). Attentional synchrony in static and dynamic scenes [Abstract]. Journal of Vision, 8(6):773, 773a.

Smith, T. J. & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision, 13*(8), 1-24.

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

Ueda, Y. & Komiya, A. (2012). Cultural Adaptation of Visual Attention: Calibration of the Oculomotor Control System in Accordance with Cultural Scenes. *PLos ONE*, 7(11), e50282.

Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition, 12*(3), 473-494.

Unsworth, N., & Engle, R.W. (2007). The Nature of Individual Differences in Working Memory Capacity: Active Maintenance in Primary Memory and Controlled Search from Secondary Memory. *Psychological Review, 114*(1), 104-132.

Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the anti-saccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology, 30*(6), 1302-1321.

Welles, O., & Zugsmith, A. (1958). Touch of Evil. In A. Zugsmith (Ed.), *Universal Pictures.*

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4

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

Chapter 2 - Executive Control of Eye-Movements in Film: Understanding How the Tyranny of Film is Attenuated by Cognitively Demanding Tasks

Attentional Selection

To understand how people comprehend and remember what they interact with in their daily lives, you must first understand what they pay attention to. Attending to information is the first step in comprehending and later remembering it. This naturally raises a key question, what influences what people selectively attend to from one moment to the next?

Through the study of attentional selection, it has been found that there are two general categories of processes that influence it: top-down and bottom-up attentional selection processes (Henderson, 2007). Bottom-up processes are guided by the stimulus features (e.g., saliency in motion, luminance, color, etc.). Conversely, top-down processes are guided based on the viewer's pre-existing cognitive state (e.g., knowledge, experience, goals). These latter processes have more recently been broken down into two distinctive types of top-down processes: volitional and mandatory processes (Baluch & Itti, 2011). Mandatory top-down processes are guided by the knowledge that the viewer has. For Instance, if I ask what time it is, you will look at your watch, cell phone, or maybe wall clock because you know that is where you generally find that information). Importantly, cultural background would be considered a mandatory process because you were raised in that with those congruent values and beliefs, if it has an impact on attentional selection during film viewing. Volitional attentional selection, the act of willfully shifting your attention, occurs when you focus on events and/or objects, related to a goal, that you would not automatically attend to. In the context of watching film, viewers automatically attend to main narrative elements (i.e., main characters, their actions, and the plot of the film). Thus, volitional attentional selection would involve deciding to attend to things that

are critical for a goal you have, especially if that goal is inconsistent with attending to the main narrative elements.

While both types of attentional selection have been studied while people look at static images, far less research has been done to investigate top-down attentional selection, mandatory or volitional, when people are looking at dynamic images (e.g., films or videos). Yet, to translate our research findings from the laboratory to the real world, investigating how people attend to films is closer to how people attend to the real world, compared to how people attend to static images. Thus, in the current study, we investigated viewers' eye-movements while they watched film clips.

Research looking at the role of top-down processes started with the use of static images, the most famous studying being done by Yarbus (1967). In this study, there was evidence for top-down processes to have a role on what participants looked at while viewing static images. Specifically, when presented with different questions to answer (e.g., "what is the SES of the people in this image" or "what is the age of the people in this image") there were distinct eye-movement patterns depending on the question that participants were asked. This suggests that there is room for the participant use their knowledge and goals to guide their attention. This general result has been replicated on several occasions when using static images (Yarbus, 1967; DeAngelus & Pelz, 2009; Smith & Mital, 2013). When the question remains that same, but the stimulus changes to dynamic scenes the results are not as clear or consistent in the research.

Research looking at the effects of top-down processes on attentional selection during film viewing has led to the creation of terms like *attentional synchrony* (e.g., looking at the same places at the same time; Dorr et al., 2010; Smith, Levin, & Cutting, 2012; Smith & Mital, 2013) and *tyranny of film* (e.g., comprehension may differ, but eye-movements do not; Loschky et al.,

2015; Hutson et al., 2017). These studies have suggested that bottom-up stimulus features of film may overwhelm top-down mandatory attentional selection. Studies that do show effects of top-down processing on attentional selection during film viewing when they engaged in a task inconsistent with comprehending the narrative, namely a volitional attention task (Lahnakoski, 2014; Hutson et al., 2017). For example, Hutson, et al., (2017) manipulated volitional control by giving participants the task of drawing a map of the film space from memory after the clip was over or to comprehend the narrative. In this condition they found large differences in attention for the two groups. The Map task condition looked less at the main narrative elements in the center of the screen, and more at the background elements, thus breaking (or attenuating) the tyranny of film. This was argued that because the Map task condition required viewers to exert volitional control, avoid looking at the main narrative elements, which made it difficult to understand the narrative. This contrasts with what viewers would automatically do when comprehending the narrative, namely attend to the main narrative elements to follow the story, as was done by viewers in the uninstructed condition.

Similarly, Lahnakoski (2014) manipulated the goals of participants by having them view a film clip as an interior designer and a detective. In these results they found that when participants were taking on the perspective of interior designer, they looked more at the background objects and less at the main characters, like participants in the map task condition in Hutson et al. (2017). In these trials, participants also reported that they had to "…actively ignore people, conversations, and/or the entire plot of the video." Conversely, when they took on the perspective of the detective, they focused on the main characters at the center of the screen in the film clip. Together, the Hutson et al. (2017) results fit nicely with these results. Specifically, when participants have a goal that is at odds with narrative comprehension, they can attenuate

the tyranny of film that is often found. Largely, this suggests that in these cases where participants had an alternative goal their attentional selection was the result of a change in the goal participants had, not just their knowledge. This suggests that it is possible to induce such volitional attentional selection while viewing film, it is unclear if this type of attentional selection is cognitively demanding for viewers. Though this can be inferred from the replies of participants when they took on the perspective of interior designer and reported they were actively inhibiting their attention to the main characters in the film clips (Lahnakoski, 2014).

Ignoring the main characters and the narrative of a film clip, to attend to background elements (e.g., the furniture, plants, decorations) may be a bit like a Stroop task (e.g., report the font color, blue, of the word "red"; Stroop, 1935). If so, then the volitional control of attention in such tasks may be more cognitively demanding than the mandatory control of attention employed to comprehend the narrative.

Culture

Cultural background has been a debated topic within the attention literature as to whether it plays a large role in attentional selection or not. Differences have been observed between western (e.g., individualistic cultures like the US, Europe, and Canada) and eastern cultures (e.g., collectivist cultures like Japan, China, and Korea; Masuda & Nisbett, 2001; Nisbett & Masuda, 2003; Chua, Boland, & Nisbett, 2005; Simonson, et al., in preparation). These studies show Western cultures focus on stimuli in a more analytic way, while Eastern cultures often focus on stimuli in a more holistic way. Ultimately, this results in Eastern cultures remembering more about how the background and foreground relate to one another (2003). These results have been found across a variety of measures such as attention, perception, memory, and comprehension.

Alternatively, there are studies that find no evidence of differences between cultures and rather show quite similar image viewing patterns. Specifically, they found no, or small, differences in fixation durations (Evans et al., 2009; Rayner et al., 2009); no differences in the number of total fixations (Evans et al., 2009; Goh, Tan, & Park, 2009; Rayner et al., 2009; Masuda, Ishii, & Kimura, 2016) or the number of fixations to background and central elements (Chua, Boland, & Nisbett, 2005; Evans et al., 2009; Goh, Tan, & Park, 2009; Masuda, Ishii, & Kimura, 2016). Rather than pointing towards differences between these cultural locations, these studies show that both Eastern and Western cultures tend to prefer the central objects of an image to the same degree. One slight variation in viewing behavior is that Eastern participants may spend relatively more time than western cultures exploring the background space, but not by a large degree. More recently it seems that the literature is leaning towards no differences in attentional selection between different cultures. This leaves the current literature to be quite divided.

An interesting point is that most culture studies focus on the use of static images, rather than dynamic images, which begs the question of whether this type of mandatory top-down attentional process would have an influence over attentional selection processes or not. However, it is possible that there would be too many strong bottom-up features, such as motion, for such attentional differences to still be present. Additionally, this study will be identifying if specific types of top-down attentional selection (e.g., goal-driven attention) are present across cultures as well.

Cognitive Load

Research has most clearly shown that volitional control of eye-movements is cognitively demanding for participants in the anti-saccade task (i.e., when looking at a computer screen, if a

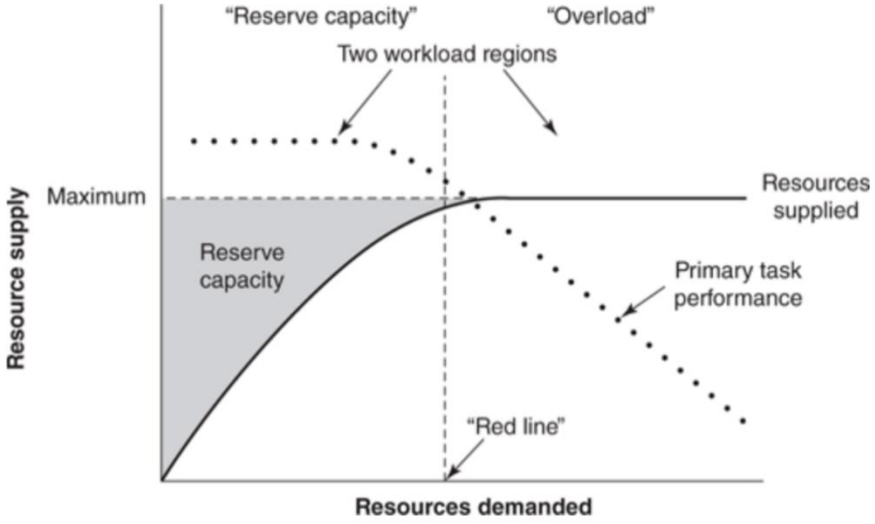
target suddenly appears on the right, look to the left; Mitchell et al., 2002; Godijn & Kramer, 2007). This task requires participants to make eye-movements in the direction of a cued stimulus (e.g., prosaccade) or in the opposite direction of the cued stimulus (e.g., antisaccade). These studies have shown that in the antisaccade task participants need to both suppress incorrect eye-movements (i.e., a reflexive eye-movement to the suddenly appearing target) and execute correct eye-movements (i.e., a volitional eye-movement in the opposite direction from the target). Additionally, those with higher working memory capacity and better executive control of attention have fewer errors (e.g., reflexive eye-movements) and perform faster overall (Unsworth, Schrock, & Engle. 2004). For these WM effects to be found though, it is imperative that the cognitive load be enough to tax the participants. Results have shown that viewers' performance on the anti-saccade task decreased (e.g., more reflexive eye-movements) when under a cognitive load (e.g., the N-back working memory task; Mitchell et al., 2002; Godijn & Kramer, 2007). Similarly, we wanted to test the hypothesis that volitional attention is cognitively demanding for viewers while watching film (e.g., when doing the Map task).

In a prior study to this paper, there was found to be no effect of cognitive load on attentional selection (Simonson et al., in preparation). This was hypothesized to be because the manipulation of cognitive load was not taxing participants enough and therefore had no impact on attentional selection. These results are not the only indication that this lack-of-an-effect of cognitive load was the potential reason we saw no effects; in fact, these ideas can be mapped onto a model proposed by Wickens et al. (2013). In Figure 2.1 you will see the model what he proposed, which shows how performance will change depending on the number of resources that are available for a person to use. In this figure, he lays out that a person will have a certain number of resources to work with (i.e., "Reserve Capacity") and that primary task performance

decreases as their reserve capacity decreases. Conversely, as they still have reserve left their performance stays quite high, because of a lack of demand that is using their available.

Figure 2.1.

Resource Supply and Resource Demanded Model Proposed by Wickens



Notes. This figure is from Wickens, et al. (2013) and depicts the proposed relationship between resources supplied, primary-task resource demand, and performance. These together indicate the work overload point, represented by the vertical dotted line.

While these ideas are related to performance on tasks related to how much WM capacity that a person has, it can also be related to the breadth of eye-movements that they would have as well (e.g., eye-movements act as a biometric measure of reserve capacity). If we use the reserve capacity to now equal the breadth of attention that they can have, we can say that as they are using more resources (i.e., "Resources supplied" line) their breadth of attention because less. This suggests that as more resources are used, they become less able to control their attentional breadth as much and over time their performance also decreases.

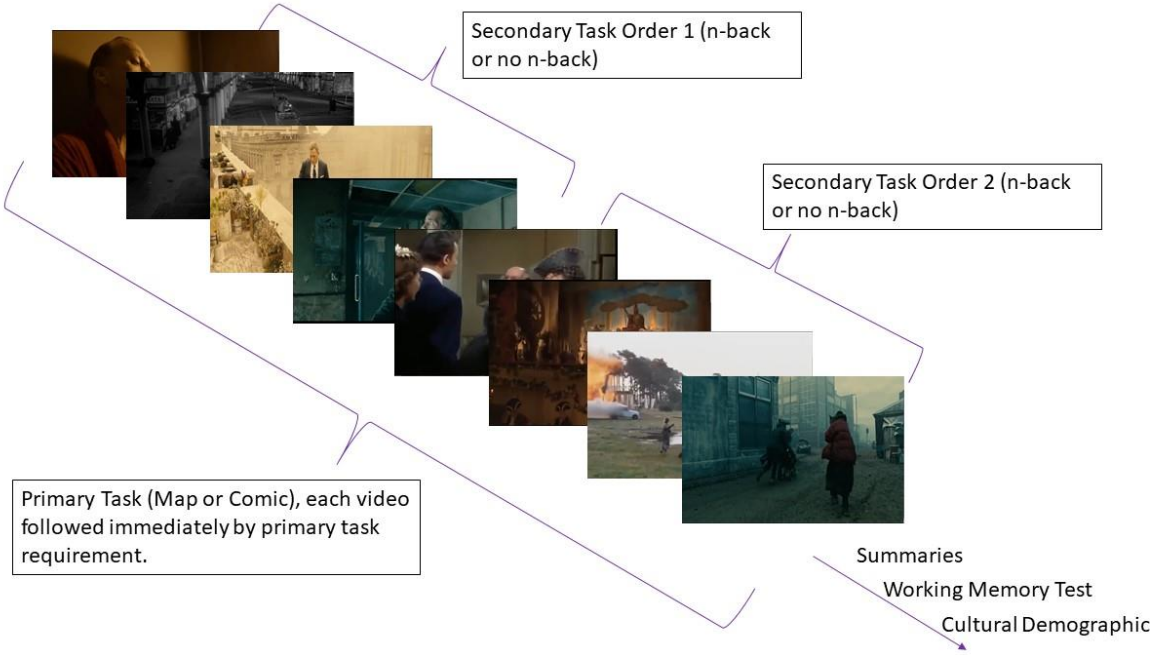
Overview of the Current Study

The purpose of this study was to answer two questions: 1) Is volitional attentional selection cognitively demanding for participants while watching film? and 2) Can cultural differences in mandatory attentional selection override bottom-up stimulus control in film?

The experimental schematic of the study is depicted in Figure 2.2 below. This study utilized 8 short film clips that participants watched while preparing for 1 of 2 primary tasks: the map task (i.e., draw a map of the landmarks (or objects) and their locations in the film space from memory after watching the film) or the comic task (i.e., draw a 4-panel comic depicting the main narrative events in the film from memory after watching the film). These primary tasks served to manipulate participants' goals while watching the film clips to manipulate their volitional attention. Additionally, on half of the trials (in either the first or second block of 4 videos), participants completed a secondary cognitive load task (i.e., Auditory 2-Back Task). After watching all videos, participants wrote summaries for all the film clips, completed an executive working memory measure, and then filled out a cultural demographic questionnaire.

Figure 2.2.

Trial Schematic for Current Study



Notes. Experiment schematic for stimuli, tasks, and conditions. 8 video clips counterbalanced across participants. The primary task (map or comic task condition) was completed after each video and was varied between-subjects. The secondary task (cognitive load or none) was done while watching videos, either during the first or second block of four videos, was varied within-subjects and counterbalanced. After all videos, participants wrote summaries for each of the videos, took an executive working memory test, and then filled out a cultural demographics survey.

Hypotheses

Top-down attention hypotheses

The *volitional attention hypothesis* predicts that participants given a goal at odds with comprehension of a narrative will have significantly different eye-movements compared to participants who were not given a task at odds with comprehension. The *mandatory attention hypothesis* predicts that the cultural background of a person will influence their attentional guidance during film viewing. While these two hypotheses are not competing, they do share a competing hypothesis. The *Tyranny of Film hypothesis* predicts that despite large differences in

cognition (e.g., based on task or cultural background) while watching the film, there will be little or no differences in eye-movements because of the strong bottom-up features of the film[5].

Cognitive load hypotheses

The *limited resources* hypothesis predicts that participants will show a decrease in their attentional breadth as their cognitive demand increases, because they only have so many attentional resources to use at any given point.

Additionally, the *executive control of eye-movements hypothesis* would predict an interaction between the presence of a cognitive load and the task that participants are completing. Specifically, when completing a cognitive load task while also completing either a goal *at odds with* comprehension of a film narrative (i.e., a volitional attention task) or a task *in line with* comprehension of the film narrative (i.e., a mandatory attention task), there should be a larger drop in attentional breadth for the former. This would suggest that the volitional attention task is more cognitively demanding compared to the mandatory attention task.

Methods

Participants

There was a total of 80 participants in this study, with participants recruited from Kansas State University (Kansas; n = 40; 16 females; average age = 19) in the United States, and Kyoto

[5] This definition of the tyranny of film hypothesis is a bit different from that used in the past. The original hypothesis was modified in this paper to accommodate the research that has been done since the original study by Loschky et al. (2015), namely Lahnakoski et al. (2014), Hutson et al. (2017), and Huff et al. (2017). The definition in the original study was not as inclusive as this new definition would imply by saying "large differences in cognition that have little or no influence on attention during film," which includes many different factors (i.e., task manipulations, cultural background, context manipulations, attitudes, etc.).

University (Kyoto; n = 40; 15 females; average age = 21) in Japan. Participants recruited from Kansas received course credit for their participation and participants from Kyoto were given bookstore credits for their participation. All participants had their vision tested using the Freiburg Acuity and Contrast Sensitivity Test (FrACT; Bach, 2007) to ensure they had normal or corrected to normal 20/20 vision and gave their informed consent before participating in the study.

Cultural demographics questionnaire

All participants completed a cultural demographic questionnaire after the experiment. We used the 2010 Census Race and Hispanic Origin Alternative Questionnaire Experiment (AQE; Compton, Bentley, Ennis, Rastogi, 2013) in our study to collect the most detailed information from our participants. This questionnaire collected information about participants' ethnicity, country of origin, and how long they had lived in the country. This enabled us to collect detailed information on each participant's cultural and ethnic background, as this was one of our independent variables.

Design

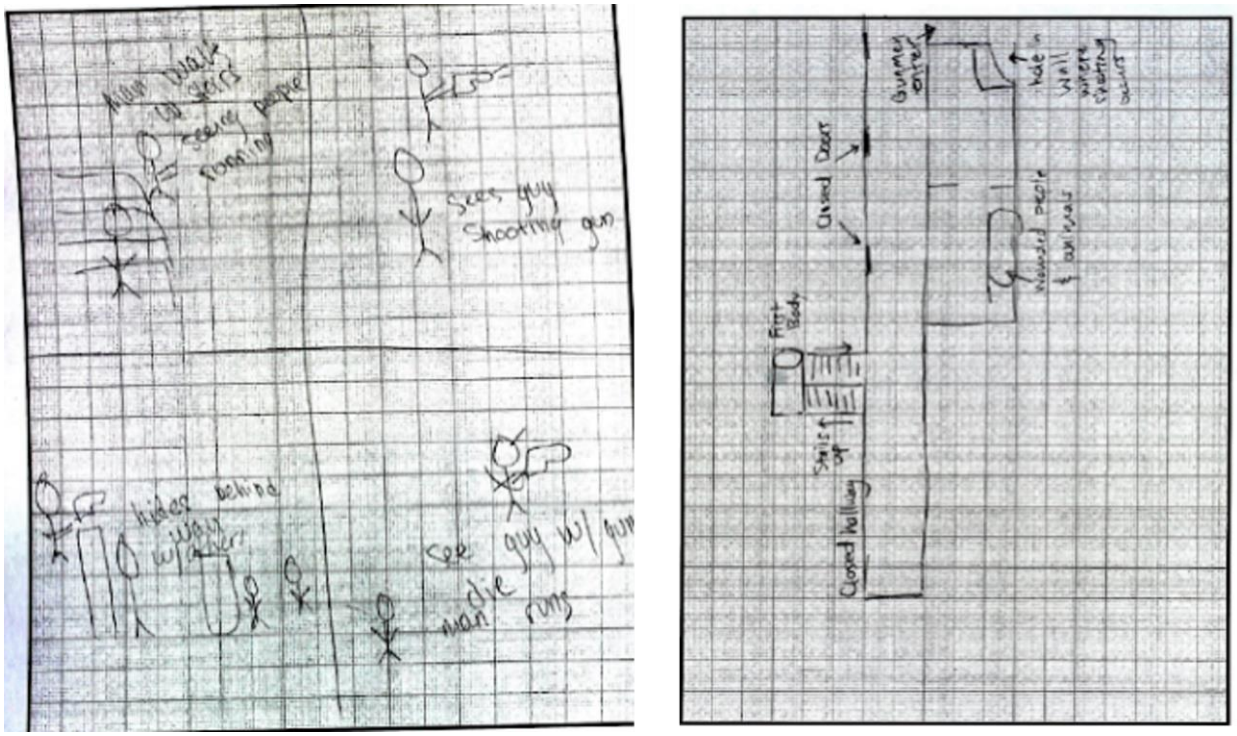
Conditions

An example of the type of responses produced by participants can be seen in Figure 2.3 below. Participants were assigned to one of two between-subject conditions: the map task condition or the comic task condition. The map task was adopted from a study by Hutson et al. (2017) and required participants to draw and label a map depicting as many objects and locations as possible from each film from memory. Participants had five minutes to complete each map, and this was our manipulation of volitional control (i.e., participants must complete the goal of drawing a map form memory). The comic drawing task condition required participants to watch

the film for comprehension of the main narrative elements occurring. In this condition, participants drew a 4-panel comic of each of the 8 experimental film clips. The comic condition served to reinforce comprehension of the main narrative elements in the film and kept participants from progressing through the experiment faster than those participants in map task condition, who had to spend time drawing maps. Importantly, the comic task served as a proper control task condition for the map task in our experiment, which was not included in the Hutson et al. (2017) study. Both tasks were created in a way to be as similar in nature as we could manage. For example, they both were required to draw pictures and word afterwards, had 5 minutes to do so, were told their task before the beginning of each clip. This ensured that any differences between the two conditions would not be to small avoidable differences in the tasks itself.

Figure 2.3.

Example Comic and Map Task Figure



Notes. Example responses for the Comic task and Map task from the same film clip (from Children of Men, Cuaron et al., 2006) [left] Example of a comic drawn in the Comic task. The participant drew the required 4-panels and described what was happening in each. [right] Example of a map that was drawn in the map task. The participant drew and labeled the objects and locations in the film space.

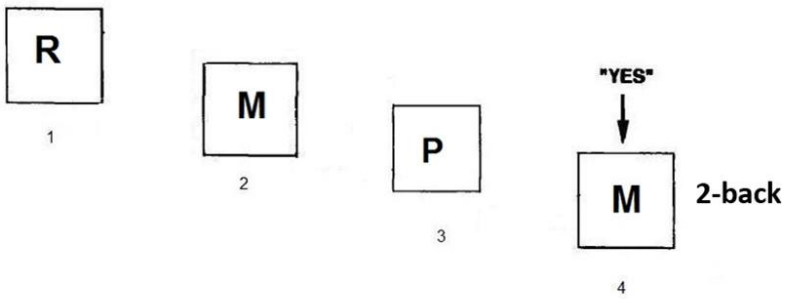
Cognitive load

The manipulation of cognitive load was a within-subjects variable. The cognitive load task that was used in this experiment is the auditory N-back task (Jaeggi et al., 2010; Ringer et al., 2014; Ringer, Throneburg, Johnson, Kramer, & Loschky, 2016). The cognitive load manipulation was completed during film viewing on half of the trials that participants completed, either during the first four film clips (1-4) or the second four (5-8) (see Figure 2.2), with block order randomly counter-balanced across participants. This task presented participants with an audio stream of letters, which were presented every two seconds during the film clip. The task

was to make a response (either ‘yes’ or ‘no’ on a Cedrus response box) within the two seconds of the auditory letter presentation. Participants needed to make a ‘yes’ response if the current letter was also present ‘N’ letters ago, and otherwise make a ‘no’ response, which required no response. A visual representation of example correct responses for the 2-back task is illustrated in Figure 2.4.

Figure 2.4.

Example of the Cognitive Load (2-back) Task



Notes. This depicts the visual analog of the auditory 2-back cognitive load task, which required participants to make a response when the letter they were auditorily presented with was the same as the letter heard 2-back.

Computers & eye-tracking

The experiment was programmed in Experiment-Builder (SR Research) and the only differences between Kansas, and Kyoto were in the language of instruction screens (English vs.

Japanese). Two senior authors (each a native speaker of English or Japanese, and highly proficient in the other language) checked all translations and made revisions as deemed necessary. All stimuli were presented on a 19” ViewSonic Graphics Series CRT monitor (Model G90fb) in Kansas and a 19.4’’ in Iiyama S103MT VisionMaster 503 CRT monitor in Kyoto.

Eye-tracking at both data collection sites was conducted using EyeLink 1000 (SR Research) eye-trackers, with default settings used, including sampling at 1000 Hz (1000 samples per second; SR Research). A chin and forehead rest were used during eye-tracking to maintain a constant viewing distance from the screen (Kansas = 25.2”; Kyoto = 22.44”). All participants went through a 9-point calibration and validation process before beginning the procedures of the study. After calibration and before beginning each video, a drift correction was used to ensure participants would maintain focus at the center of the screen and that the calibration was accurate throughout the study. Specifically, as per SR Research guidelines, a maximum error of 1° and a maximum average error of 0.5° visual angle were obtained for all calibrations and maintained throughout the study.

Film clips

Eleven different film clips were used in the experiment. The clips were between 1:00 and 3:30 minutes in length and were all unedited, Hollywood style long-shots[6]. The clips were presented at a rate of 30fps and a resolution of 1280 X 1024 pixels at both Kansas (screen size = 14.38’’ X 10.75” = 36.53 cm X 27.31 cm) and Kyoto (screen size = 15.5’’ X 11.6” = 39.37cm X

4. [6] The reason for using only long-shots from these films is that Hollywood-style editing typically uses short shots that, on average, occur every 2-4 seconds. This has been shown to strongly affect eye-movements, because viewers almost invariably move their eyes to the center of the screen immediately after a new shot appears, thus artifactually increasing viewers’ Attentional Synchrony (Dorr, et al., 2010; Smith, Levin, & Cutting, 2012; Smith & Mital, 2013).

29.46 cm). The screen subtended 31.48° X 24.08° of visual angle in Kansas and 36.11° X 28.98° of visual angle in Kyoto. Clips were chosen based on two criteria: 1) there needed to be enough background landmarks, objects, and different locations to ensure that participants in the map task condition remained engaged with them, and 2) there needed to be enough main narrative elements in the film clip to ensure that participants in the comic task condition remained engaged with them. Of these 11 clips, three were used for practice trials, to ensure that the participants were familiar with and able to perform their respective tasks. The source films were Birdman (from "Birdman," Iñárritu et al., 2014), Touch of Evil (from "Touch of Evil," Welles et al., 1958) , James Bond: Spectre (from "James Bond: Spectre", Mendes et al., 2015), Children of Men (2 clips; from "Children of Men," Cuaron et al., 2006), Rope (from "Rope," Hitchcock et al., 1948), The Russian Ark (from "The Russian Ark," Sokurov et al., 2002), and Sacrifice (from "The Sacrifice," Tarkovsky et al., 1986). Details of each of the eight experimental clips are given in the Supplemental Materials. The order in which each video was presented used a Latin Square design across tasks and participants for each culture.

Executive working memory measure

Participants performed the Operation Span (OSPAN) task, which is designed to test the executive-working memory (E-WM) capacity of participants (Bailey, 2012; Conway et al., 2005). The OSPAN task required participants to remember words while they mentally evaluated the truth of math equations. Participants had 4 seconds to determine whether a math equation was true or false. For example, if a participant saw "(10/2) + 3 = 7," they should have responded "incorrect." After participants responded to each math problem, a word was flashed on the screen for 1 second. Participants completed a random number of math equation trials (between 3-7)

within a block and were then instructed to recall as many words as they could from within the block (15 blocks total).

Cultural controls

The same experiments were run at both Kansas and Kyoto. All aspects of the study were discussed in detail, to ensure that the two studies were identical to one another between the two laboratories. All the instructions were translated by graduate students at Kyoto who speak both English and Kanji and were checked by professors at Kansas and Kyoto who speak both languages well, to keep the meaning and phrasing as close as possible within both languages.

Procedure

Participants were randomly assigned to 1 of 2 between-subject task conditions (map or comic task) before entering the lab. After giving their informed consent, participants had their vision tested. They then went through eye-movement calibration and validation for the EyeLink 1000. After calibration, participants watched the three practice videos. The first video was used to practice the task for the condition they were randomly assigned (map or comic task). The second practice video was to have participants practice the secondary auditory cognitive load task by itself (more details about the secondary auditory cognitive load task are given in the Supplementary Materials). The third video was to have participants practice the dual-task paradigm (i.e., the map or comic task together with the cognitive load task).

A visual schematic of the main experiment can be seen in Figure 2.2 above. After the practice film clips, participants completed the 8 experimental trials. In each of those, they watched a video in preparation to draw and write either a map or a 4-panel comic from memory afterwards. On half of the trials, either the together with video clips 1-4 or clips 5-8, participants

completed a second within-subjects manipulation (N-back cognitive load) with 2 levels [cognitive load present or absent] for the duration of the film clip.

After watching all the film clips, participants wrote free recall summaries of the main narrative elements that occurred within each of the clips. Participants were prompted with two frames from each of the film clips, one from the first second of the film clip and one from the last second of the film clip. This had two purposes: 1) to serve as retrieval cues for the correct video-related information, 2) to minimize the amount of information the participant was given as retrieval cues, to avoid giving away map and comic task-related information. Finally, participants then completed the OSPAN working memory task and the cultural demographic questionnaire. After completing all the above, participants were debriefed and thanked for their participation.

Results

Organization of results

We will first discuss the results in terms of our predictions. After this discussion, we will show results from exploratory analyses for different performance measure variables (e.g., N-back measures, Map measures, Task trade-offs). Lastly, we will investigate the role of several different variables, such as time, working memory, and general performance.

Dependent measures

There are two main outcome measures related to eye-movements that we will discuss: Deviation from Screen Center and Narrative-AOI fixation counts. Deviation from screen center is understood to measure the viewer's breadth of attention Findlay & Walker, 1999), which is important for testing both the *Volitional Attention* and *Mandatory Attention* hypotheses. Narrative-AOI (areas of interest) fixation counts can be understood to measure the amount of

attention paid to the central narrative elements in each film clip. Thus, it is also important for testing the *Volitional Attention hypothesis*, which predicts more attention to those narrative elements by participants in the Comic task than those in the Map task (Hutson et al., 2017; Lahnakoski et al., 2014). Likewise, narrative-AOI fixation counts are important for testing the *Mandatory Attention hypotheses*, which predicts more attention to the central narrative-AOIs by those participants from a Western culture than an East Asian culture (Masuda & Nisbett, 2001; Nisbett & Masuda, 2003; Chua, Boland, & Nisbett, 2005; Goh, Tan, & Park, 2009).

These measures were chosen based on research from Findlay and Walker (1999) that argue saccade amplitudes are measuring the "where" of eye-movements, while fixation durations measure the "when" of eye-movements; importantly, emphasizing that these are separate, distinguishable processes. Here we want to focus on the "where" system to address attentional selection between our different independent variables (e.g., location, task, cognitive load condition). For this reason, we focused on deviation from screen center and the complementary measure of the number of fixations in the central narrative-AOIs, to measure attentional selection, rather than eye movement measures concerned more with the "when" system, such as fixation durations, or saccadic latencies. Ideally, the deviation from center and number of fixations in the central AOI measures should produce inverted relationships (e.g., when deviation from screen center goes up, the number of fixations in the central narrative-AOIs should go down)[7]. Having established that, we will then focus only on the deviation from screen center results in our exploratory analyses to streamline the discussion.

[7] Measures used in this study were additionally based on the results of correlations between a larger set of variables used in a prior study (e.g., Fixation Durations, Proportion of Saccades, Proportion of Deviation, Narrative-AOI Dwell Time, Narrative-AOI Fixation Count, and

Data preparation

Eye-movement data were cleaned to remove outliers according to the normal procedures for fixation durations in the field. Any fixation durations above 3,000 ms (3 seconds) or below 40 ms were excluded from the analyses. The measure of deviation from screen center used the fixation x and y locations relative to the computer screen center and was measured in degrees of visual angle. This measure 1) used the size of the screen diagonal in degrees to convert from pixels to a measure in degrees and then 2) to make comparisons across cultures, the degree measure was divided by half the screen diagonal (in degrees) to measure the proportion of the maximum deviation from screen center. Narrative areas of interest (narrative-AOIs) were created and used for two separate outcome measures: Dwell Time in narrative-AOIs and number of fixations in the narrative-AOI. These narrative-AOIs were placed on content that was considered maximally relevant to following the narrative of each film clip, as judged by both the lead author, and confirmed by several research assistants.

Model creation

Most data analyses were performed using R statistical software (version 3.1.1), any differences in this will be outlined and discussed in the supplementary materials. We used the lme4 (Bates, Mächler, Bolker, & Walker, 2014) library to run our mixed models, the emmeans

---

Proportion of Fixations to Narrative-AOIs; Simonson, et al., in prep). For the current study, we chose variables that were most theoretically relevant from that larger set from the theoretically relevant to be used in this paper. The correlations between the excluded variables and the two measures included here (e.g., Deviation from screen center and number of fixations to narrative AOIs) can be found in supplementary materials. previous study. The correlations between the excluded variables and the two measures included here (e.g., Deviation from screen center and number of fixations to narrative AOIs) can be found in supplementary materials.

(Lenth, 2018) library to plot least squares predicted means from the model fit, the afex (Singmann, et al., 2015) library to get parameter specific *p* values from the models, and the ggplot2 (Wickham, 2016) library for figure creation. To determine the best fit models AIC (Akaike, 1973) and BIC (Schwarz, 1978) values were used, since BIC is more conservative this value was favored in the comparisons. Of note, the temporal order of the auditory N-back task (i.e., while watching video clips 1-4, vs. clips 5-8), while controlling for all other predictors, was not a significant predictor of the proportion of deviation ($p = .769$), Narrative-AOI fixation count ($p = .823$), N-back Performance ($p = .182$), or Map Performance ($p = .779$) so it was not included in any models through the rest of the paper.

Model results: deviation from screen center

We ran a gamma generalized multilevel model with a square root link function to determine if proportion of deviation from screen center differed between culture, task condition, or cognitive load levels. The best fit model included full-factorial effects of Location [Kyoto= -1, Kansas = 1], Task Condition [Comic = 1, Map = -1] and Cognitive Load [Absent = -1, Present = 1] as well as video and subject as random effects at their intercept.

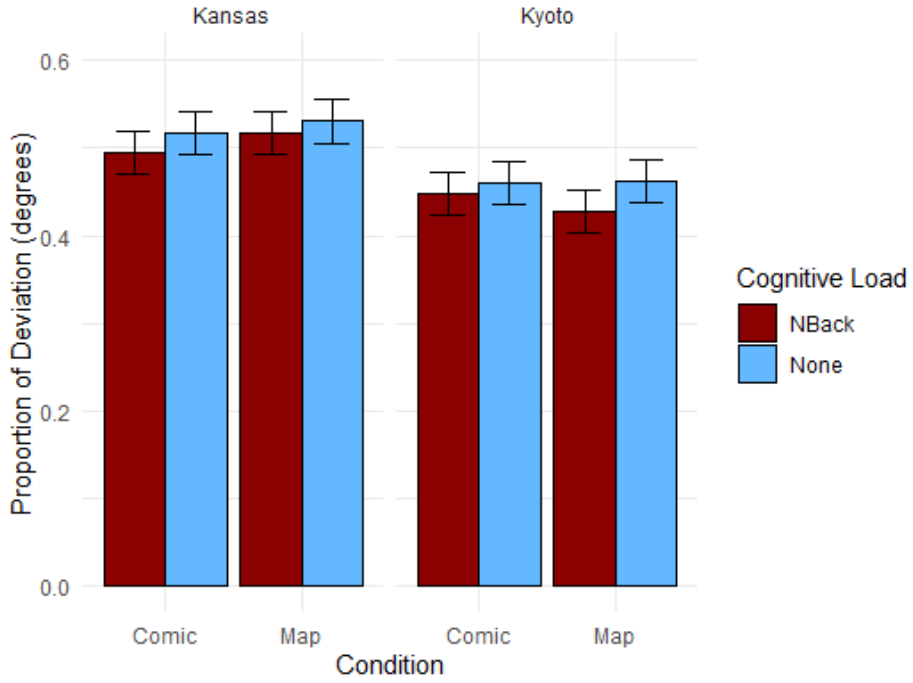
The results of the model are depicted visually in Figure 2.5 and the results can be found in more detail in Table 2.1 below. The main effect of task condition was not a significant predictor of deviation from screen center, which refuted the *volitional attention hypothesis.* Consistent with the *limited resources hypothesis*, there was a main effect of cognitive load on deviation; there was less deviation on trials in which participants were under the dual-task 2-back cognitive load ($M = 0.47$, $SE = 0.01$) compared to the single task trials without the 2-back cognitive load ($M = 0.49$, $SE = 0.01$). Consistent with the *mandatory attention hypothesis* the main effect for location was significant, such that the Kansas participants ($M = 0.51$, $SE = 0.01$)

showed more deviation from screen center compared to the Kyoto participants (M = 0.45, SE = 0.01).

There was a significant Condition Type × Location interaction, indicating that the effect of task varied by culture, specifically the lowest deviation was for Kyoto participants in the Comic task Condition. There was also a Cognitive Load × Location interaction, such that Kyoto participants showed greater effects of the N-back cognitive load on their deviation from screen center than the Kansas participants. Furthermore, in support of the *executive control of eye-movements hypothesis*, the Condition Type × Cognitive Load interaction was significant. This effect is more fully reflected in the significant 3-way interaction of Condition Type × Cognitive Load × Location. Specifically, the significant three-way interaction shows that participants from Kyoto had a greater decreased deviation, when under a cognitive load and in the map condition, compared to the Kansas participants. Put differently, the Kyoto participants showed the predicted 2-way interaction between task and cognitive load, in support of the *executive control of eye-movements hypothesis*, but the Kansas participants did not.

Figure 2.5.

Results of Generalized Multilevel Model Predicting Proportion of Deviation



Notes. Results of a generalized multilevel model where the outcome variable was the proportion of maximum deviation from screen center (degrees of visual angle). The values here represent the back-transformed proportional data and error bars are 95% confidence intervals.

Table 2.1.

Proportion of Deviation Results

Parameter estimates of a full-factorial generalized multilevel model predicting the proportion of deviation from screen center (degrees of visual angle).

Term	Estimate	*SE*	*t*	*p*
Intercept	0.482	0.012	39.778	<.001*
Condition Type [Comic]]	-0.002	0.003	-0.758	.448
Cognitive Load [Present	-0.106	3.6e-4	-29.274	<.001*
Data Collection Location [Kansas]	0.032	6.5e-4	49.560	<.001*
Condition Type [Comic] × Cognitive Load [Present]	0.002	3.6e-4	4.545	<.001*
Condition Type [Comic] × Location [Kansas]	-0.007	6.4e-4	-10.894	<.001*
Cognitive Load [Present] × Location [Kansas]	0.002	3.6e-4	4.612	<.001*
Condition Type [Comic] × Cognitive Load [Present] × Location [Kansas]	-0.004	3.6e-4	-11.019	<.001*

*Note: * are significant at the p<.05 level. Variables are effect-coded, and all referenced to the overall average deviation from screen center (intercept).*

Model results: narrative-AOI fixation count

We ran a Poisson generalized multilevel model with a log link function to determine if the number of fixations made to the narrative-AOIs differed between culture, task condition, or cognitive load levels. The best fit model included full-factorial effects of Location [Kyoto= -1, Kansas = 1], Task Condition [Comic = 1, Map = -1] and Cognitive Load [Absent = -1, Present = 1] as well as video and subject as random effects at their intercept.

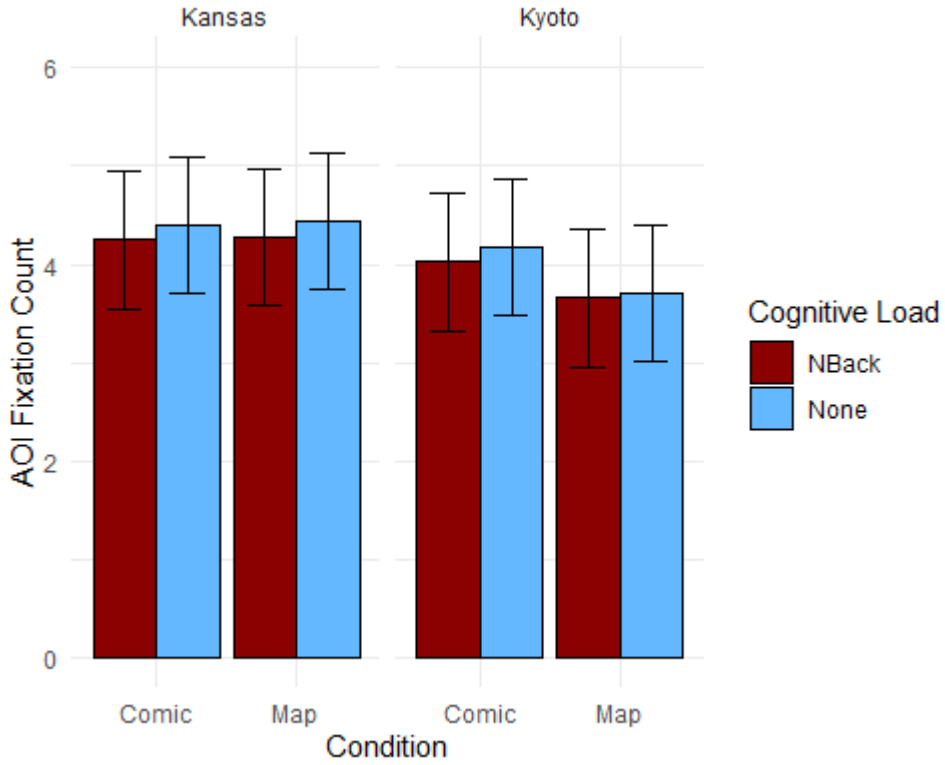
The results of the model are depicted visually in Figure 2.6 and the results can be found in more detail in Table 2.2 below. In support of the *volitional attention hypothesis*, there was a significant effect of task condition; fewer fixations were made to the narrative-AOIs when in map task (M = 4.02, SE = 0.35) compared to the comic task (M = 4.21, SE = 0.35). Consistent with the *limited resources hypothesis*, there was a main effect of cognitive load on fixation count; there were fewer fixations to narrative-AOIs on trials where the 2-back cognitive load was present (M = 4.05, SE = 0.35) compared to the single task trials where the 2-back cognitive load was absent (M = 4.18, SE = 0.35). Consistent with the *mandatory attention hypothesis* the main effect for location was significant, such that the Kansas participants (M = 4.34, SE = 0.35) had more fixations to narrative-AOIs compared to the Kyoto participants (M = 3.89, SE = 0.35).

There was a significant Condition Type × Location interaction, indicating that the effect of task varied by culture, specifically the fewest fixations in narrative-AOIs occurred for Kyoto participants in the Map task condition. There was a significant effect of the Cognitive Load × Location interaction, specifically the fewest fixations in the narrative-AOIs occurred for Kyoto participants when the 2-back cognitive load was present. These effects are in support of the *mandatory attention hypothesis* and the *executive control of eye-movements hypothesis,* respectively. There was no significant effect of the Condition X Cognitive Load interaction, but

this is more fully reflected in the significant 3-way interaction of Condition Type × Cognitive Load × Location. Specifically, the significant three-way interaction shows that generally the presence of cognitive load decreased fixations to narrative-AOI, but this is not true for Kyoto participants when in the dual map and N-back trials where there are an equal number of narrative-AOI fixations being made.

Figure 2.6.

Results of Generalized Multilevel Model Predicting AOI Fixation Count



Notes. Results of a generalized multilevel model where the outcome variable was the Narrative-AOI Fixation Counts. The values here represent the raw, back-transformed data, and error bars are 95% confidence intervals.

Table 2.2.

Narrative-AOI Fixation Count Results

Parameter estimates of a full-factorial generalized multilevel model predicting Narrative-AOI Fixation Count.

Term	Estimate	*SE*	*z*	*p*
Intercept	4.11	0.352	11.698	<.001*
Condition Type [Map]	0.096	0.041	2.352	.019*
Cognitive Load [Absent]	-0.062	0.007	-8.85	<.001*
Data Collection Location [Kyoto]	0.224	0.012	20.803	<.001*
Condition Type [Map] × Cognitive Load [Absent]	-0.012	0.007	-1.68	.093
Condition Type [Map] × Location [Kyoto]	-0.112	0.011	-10.382	<.001*
Cognitive Load [None] × Location [Kyoto]	-0.014	0.007	-1.969	.049*
Condition Type [Map] × Cognitive Load [Absent] × Location [Kyoto]	0.012	0.007	1.753	.079

*Note: * are significant at the p<.05 level. Variables are effect-coded, and all referenced to the overall average deviation from screen center (intercept).*

Discussion

The results provide support for the *mandatory attention hypothesis*. This is important as the existence of the cultural effect on attentional selection shown here has been a highly contentious issue (Masuda & Nisbett, 2001; Nisbett & Masuda, 2003; Chua, Boland, & Nisbett, 2005; Goh, Tan, & Park, 2009; Rayner et al., 2007; Rayner, Castelhano, & Yang, 2009; Ueda & Komiya, 2012; Senzaki, Masuda & Ishii, 2014; Masuda, Ishii, & Kimura, 2016). This result also replicates and extends results from past studies from our laboratories (Simonson et al., in preparation). Specifically, Kyoto participants' eye-movements were less exploratory but spent less time within the Narrative-AOIs overall, suggesting they took in more of the information outside those Narrative-AOIs on the screen in the film clips, across task and cognitive load conditions. Replicating our previous findings (Simonson et al., in preparation), adds greater strength to them, and provides further support for the existence of culturally-based differences in viewers' overt attention. The results also replicate our previous finding that such cultural differences in viewers' eye-movements can be found even while watching film, showing further conditions in which, the tyranny of film can be attenuated. Importantly, these cultural results would be classified as a mandatory top-down attentional process and show that not only volitional attention can break the tyranny of film.

Additionally, these results show support for the *executive control of eye-movements hypothesis* because there was a significant Condition X Cognitive Load interaction and a significant Condition X Cognitive Load X Location, 3-way interaction. Specifically, Kyoto participants in the dual-task combination of Map task + N-back showed a greater decrement in their eye-movement control (i.e., even less deviation from center) than in the dual-task combination of Comic task + N-back. This suggests that the Map task was more cognitively

demanding than the Comic task, since eye-movements become more focal in those dual-task trials. Importantly, however, these conclusions can only be drawn for the Kyoto participants, because we found no evidence of the same effect for Kansas participants. Surprisingly, there was a similar but smaller, and non-significant effect for Kansas participants when in the Comic task + N-back trials, which might suggest that participants were more challenged in these circumstances; this was not the predicted effect because the comic task was believed to be more automatic than the map task.

This raises the question, why was the *executive control of eye-movements hypothesis* only supported for Kyoto participants, but not the Kansas participants? Thus, our focus shifted to explaining this discrepancy. We developed two relatively simple hypotheses for explaining the lack of the predicted interaction between task and cognitive load for the Kansas participants:

- The *greater resources hypothesis* predicts that Kansas participants had greater cognitive resources than the Kyoto participants, and thus were not subject to the predicted decrement in volitional attention control under the dual Map task + N-back cognitive load.
- The *reserving resources hypothesis* predicts that Kansas participants did *not* have greater cognitive resources than the Kyoto participants. Rather, it predicts that the Kansas participants used dual-task performance trade-offs to maintain their resource reserve capacity, and thus reduced their cognitive load. Therefore, they did not show the predicted decrement in volitional attention control under the dual Map-task + N-back cognitive load.

To test these alternative competing hypotheses, we carried out exploratory analyses, which we report below. For simplicity we will report on the deviation from screen center

measure of eye-movements since this is where the unexpected initial 3-way interaction was found.

Exploratory Analyses: N-back

To test the *greater resources hypothesis*, we compared N-back performance between the Kyoto and Kansas participants, to see if the Kansas participants N-back scores were higher than, or at least equal to, the Kyoto participants, particularly when doing the Map task. Conversely, the *reserving resources hypothesis* would predict that the Kansas participants would have significantly lower N-back scores than the Kyoto participants, particularly when performing the Map task.

Data preparation

Eye-movement measures were prepared in the same ways as discussed above. N-back performance was used in two different ways. First, average N-back performance was measured for each participant on each trial. There were two data points that were eliminated for being below 50% accuracy (e.g., chance performance). Second, N-back performance was measured on each N-back letter presentation. In this method participant performance was treated as either "correct" or "incorrect" on each presentation (every 2 seconds).

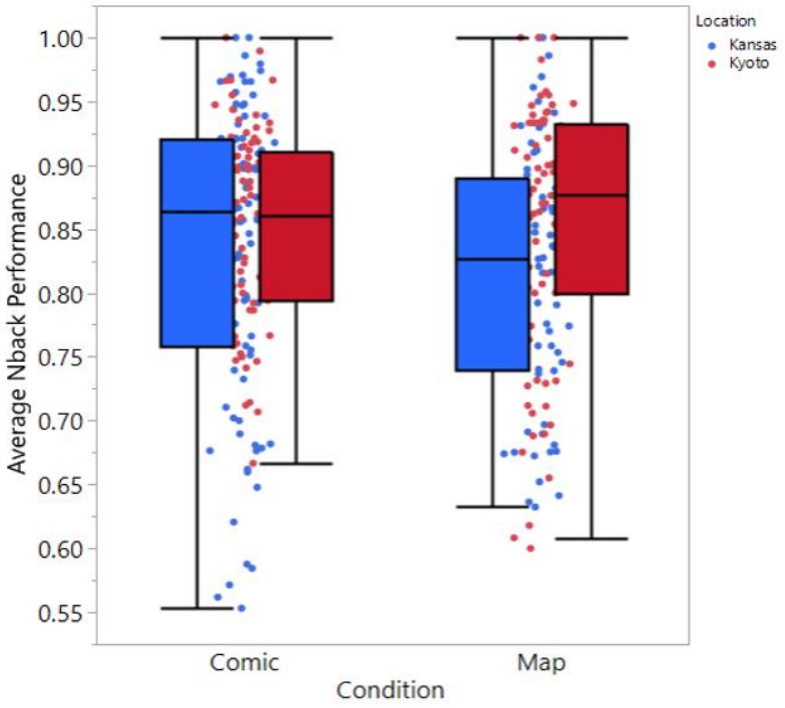
N-back performance analysis

Using a linear multilevel model with Location [Kanas = 1, Kyoto = 0] and Condition [Comic = 1, Map = 0] entered as fixed effects and video and subject entered as random effects, Average N-back performance [range: 0.55 – 1.0] was predicted. N-back performance here is a single averaged value per participant on each of their four dual-task trials (4 data points per person).

Results are shown in Figure 2.7 and Table 2.3 below, which indicate an effect of Location; Kyoto participants (M = 0.86, SE = 0.01) performed better on the N-back cognitive load task than Kansas participants (M = 0.81, SE = 0.01). This is *in*consistent with the *greater resources hypothesis* that predicted Kansas participants had greater cognitive resources than Kyoto participants. However, Kansas participants' lower N-back performance is not sufficiently informative regarding the *reserving resources hypothesis*. For that, the effect of task condition (Map vs. Comic) on N-back performance is necessary. Specifically, if Kansas participants traded-off N-back performance for Map task condition performance, we would expect them to have worse N-back performance in the Map task condition than the Comic task condition. However, the analysis showed no significant effect of Condition, which is inconsistent with the *reserving resources hypothesis*. Specifically, this suggests that neither Kyoto nor Kansas participants had a harder time completing the cognitive load N-back task in either the Map or Comic task conditions.

Figure 2.7.

Results of Multilevel Model Predicting N-Back Performance



Notes. Results of a multilevel model where the outcome variable was Average N-Back Performance. The values here represent the raw, untransformed data and error bars are 95% confidence intervals.

Table 2.3.

Average N-back Performance

Term	Estimate	*SE*	*t*	*p*
Intercept	0.833	0.012	67.26	<.001*
Location [Kansas]	-0.025	0.004	-5.59	<.001
Condition [Comic]	0.009	0.012	0.77	.446

*Note: Values denoted by * are significant at the p<.05 level.*

We thought that our above analyses might have missed more subtle variations in N-back performance over time. For example, a trade-off between N-back performance and Map task performance, which only took place later in trials, might produce the somewhat lower N-back

scores in Kansas, and the lack of the expected Task x Cognitive Load interaction in Kansas, but might not be enough to create a main effect of task on N-back performance. Thus, we next completed an analysis that looked at these potential differences in N-back performance over time.

N-back & time analysis

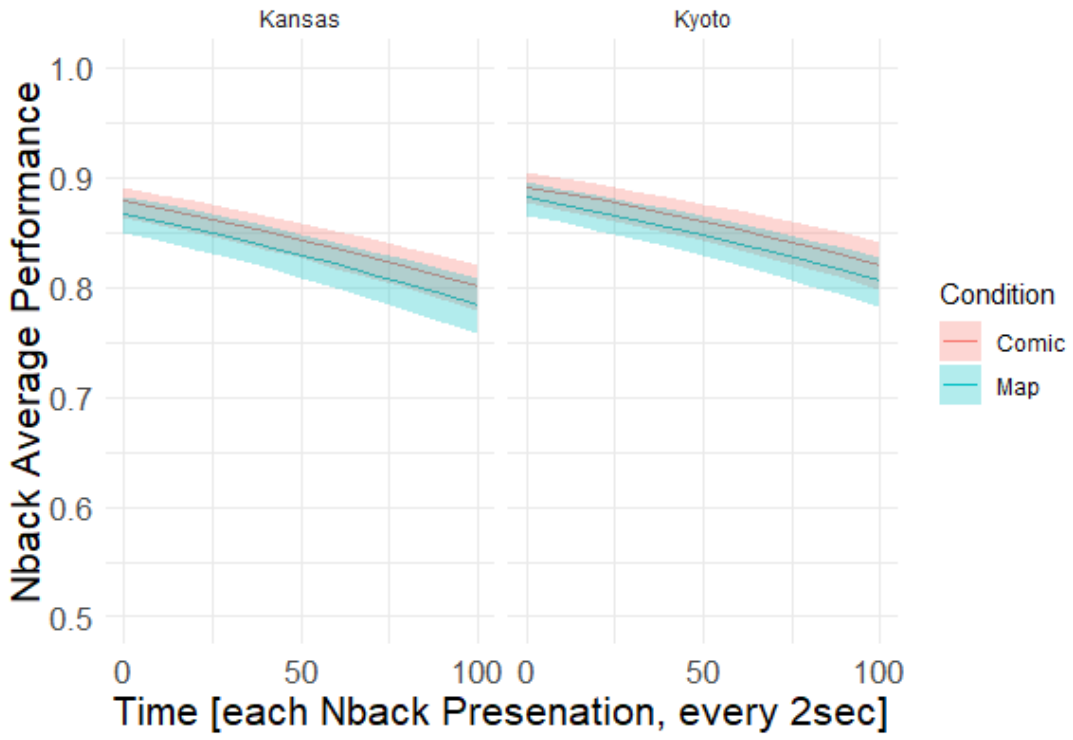
A logistic generalized multilevel model analysis was run to determine if N-back performance was changing over time. Here N-back Performance is being operationalized on a trial-by-trial basis, where participants receive either a 1 if they answered correctly or a 0 if they answered incorrectly. The measure of time that is being used here is a measure of N-back presentation. Specifically, the N-back was presented every two seconds and that constant presentation of N-back auditory letters is what is being used to represent time in this analysis. Model comparisons were made to determine the best fit random and fixed effects for the data, these comparisons can be found in the supplementary materials. The best fit model had main effects of Condition [Map =1, Comic = 0], Location [Kyoto = 1, Kansas = 0], and Time [Range: 1 - 97] as well as video and subject as random effects at their intercept.

Figure 2.8 and Table 2.4 show the results of this model. There were no significant effects of Location or Condition, thus providing no support for the *reserving resources hypothesis* (Our change in how we operationalized N-back Performance, trial-by-trial basis, rather than in terms of mean accuracy, could be why the location effect was not significant in this analysis.). However, there was a significant effect of Time when predicting N-back performance, such that both Kyoto and Kansas participants' N-back performance decreased over time. This suggests that the N-back task was taxing participants' cognitive resources over time, since otherwise, we

might have expected their performance to increase with time (a practice effect), or at least stay the same.

Figure 2.8.

Results of Generalized Multilevel Model Predicting N-back Performance



Notes. Results of generalized multilevel model where the outcome variable was Average N-Back Performance. The values here represent the model predictions and error bars are 95% confidence intervals.

Table 2.4.

Average N-Back over Time Results

Parameter estimates Condition, Location, and Time predicting N-Back Performance.

Term	Estimate	*SE*	Z	*p*
Intercept	1.976	0.126	15.741	<.001*
Condition Type [Map]	0.133	0.155	0.855	.392
Location [Kyoto]	-0.1	0.155	-0.639	.523
Time	-0.006	0.001	-7.646	<.001*

*Note: Values denoted by * are significant at the p<.05 level.*

There were additional analyses (OSPAN analysis) and hypotheses created for the N-back focused results that can be found in the Supplementary Materials. These analyses were not included in the main text of the paper for brevity.

Discussion

The N-back results showed that Kyoto participants had a higher overall N-back performance than Kansas participants, which was inconsistent with the *greater resources hypothesis*. However, the results showed no difference in N-back performance across the Map and Comic task conditions for either Kyoto or Kansas participants, which was also inconsistent with the *reserving resources hypothesis* for Kansas participants. The difference in N-back performance between Kansas and Kyoto was significant, but relatively small (81% vs. 86%), which suggests something else accounted for the large differences in attentional breadth found in our initial 3-way interaction (see Figure 2.5).

We further tested if such a trade-off might have been limited in time, such as later in each video. We did find that participants performed worse on the N-back task over time, consistent with the idea that it was a cognitively taxing task. This is an important finding because we can look at the impact of N-back in real-time, rather than just as a single averaged number, which

tells us that participant behavior changes throughout the experiment. However, there was no difference between Kyoto and Kansas, or across the Map and Comic tasks. Importantly, this is a more sensitive measure of N-back performance compared to overall averages and it solidified our belief that the N-back difference between Kansas and Kyoto was not a large enough to explain the unexpected initial 3-way interaction. Thus, again, we found no support for the *reserving resources hypothesis*. This suggests that participants in both Kyoto and Kansas gave high priority to their performance of the N-back task.

Therefore, to further test the *reserving resources hypothesis*, we asked whether the Kansas or Kyoto participants might not have performed the map task correctly, and therefore left more resources for accurate completion of the 2-back cognitive load task. Conversely, it was still possible that Map performance would show results consistent with the *greater resources hypothesis*, if the Kansas participants had significantly higher Map task scores than the Kyoto participants. These questions led to our next set of exploratory analyses, which investigated the map task performance.

Exploratory Analyses: Map Task Performance

Data preparation

Map performance had a single value per video, thus eight values per participant. We did not exclude any map scores from the analyses, as low performers were equally as important as those high performers in the following analyses.

Map scoring procedures

To score the hand-drawn maps, we used the Gardony map drawing analyzer (GMDA) (Gardony, Taylor, & Brunyé, 2016). The GMDA software allows one to compare the accuracy of depicted landmarks or objects in a participant's map, as well as the accuracy of the depicted

spatial relationships between them. To do this, we created a Master Map that represented the ground truth of what was shown in each film clip. The creation of these master maps is discussed in more detail in the supplementary materials. We scanned the Master Maps into the GMDA software and used them as templates for scoring the accuracy of participants' maps. Each participant map was hand-coded relative to the Master Map by two trained undergraduate research assistants who placed landmark/object markers on the corresponding locations within each scanned participant map. Then the GMDA software compared both the number of landmarks/objects in a participant's map, and their relative spatial locations, with the master map, and computed an overall accuracy score. Average similarity was recorded for our coders and can be found in the Supplementary Materials. More detailed discussion of how the maps were scored is given in the Supplementary Materials.

Map performance

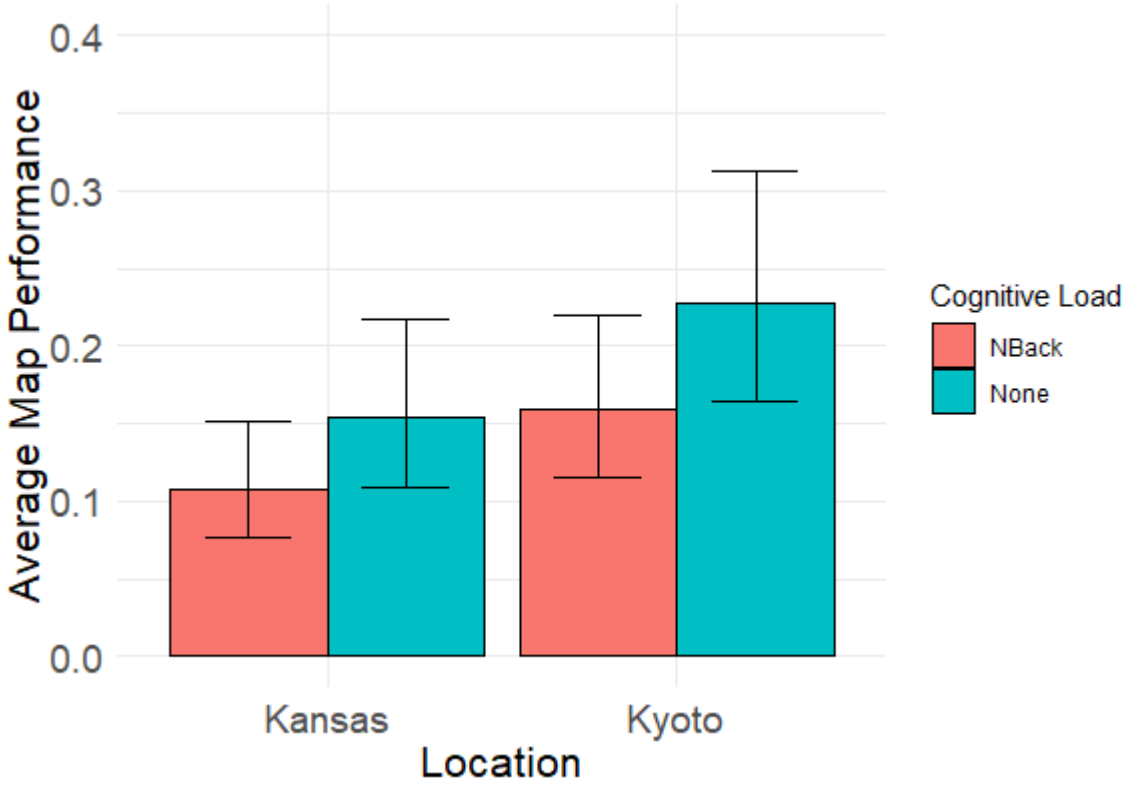
We ran a gamma generalized multilevel model with a log link function to determine if map performance differed between culture or cognitive load levels. We made model comparisons to determine the best fit random and fixed effects for the data, and these comparisons can be found in the Supplementary Materials. The best fit model had main effects of Location [Kyoto= 1, Kansas = 0] and Cognitive Load [Absent = 1, Present = 0] as well as video, subject and trial as random effects at their intercept.

Results are shown in Figure 2.9 and Table 2.5 below and show support for the *reserving resources hypothesis*. Specifically, Map task performance was significantly predicted by cognitive load, such that cognitive load being present (M =-2.04, SD = 0.16) produced worse Map task performance than cognitive load being absent (M = -1.68, SD = 0.16), and participants

from Kansas (*M* = -2.05, *SD* = 0.17) showed significantly lower Map task scores than those from Kyoto (*M* = -1.66, *SD* = 0.16).

Figure 2.9.

Results of Generalized Multilevel Model Predicting Average Map Performance



Notes. Results of generalized multilevel model where the outcome variable was N-Back Performance. The values here represent the raw, untransformed data points and error bars are 95% confidence intervals.

Table 2.5.

Average Map Performance Results

Parameter estimates Cognitive Load and Location predicting Map Performance.

Term	Estimate	*SE*	*t*	*p*
Intercept	-2.233	0.175	-12.777	<.001*
Cognitive Load [Absent]	0.359	0.094	3.829	<.001*
Location [Kyoto]	0.393	0.088	4.475	<.001*

*Note: Values denoted by * are significant at the p<.05 level.*

Discussion

The Map task results showed that participants traded off Map task performance for N-back performance. This was true for both Kyoto and Kansas participants, but the Kansas participants showed significantly larger trade-offs. The only remaining question that we had was if the task trade-off was different at our two locations. Since there was not a large performance difference between Kansas and Kyoto on the cognitive load manipulation (i.e., N-back accuracy of 81% vs 86% respectively), but a larger difference in the map task (i.e., Map task accuracy of 11.75% vs 21.35% respectively), it is possible that Kansas participants traded off between the two tasks and prioritized the N-back over the Map task. Conversely, it looks like Kyoto participants engaged in less task trade-offs and prioritized both tasks more equally highly. We therefore investigated the potential role of the trade-offs between the primary and secondary tasks in determining the 3-way interaction between Task, Cognitive Load, and Culture on participants deviation from screen center. To do that, we focused on the deviation from center in the Map task under the N-back cognitive load for both Kansas and Kyoto. For that, we investigated the relationships between Map task accuracy and N-back accuracy on deviation from center.

Exploratory Analyses: Task Trade-offs

Several models were compared to determine if a model that included a 2-way interaction was a better fit than just a main effects model. These comparisons can be found in the Supplementary Materials. The best model of the data included Location [Kyoto = 1; Kansas = 0], N-back Performance [Range: 0.55 - 1], and Map Performance [Range: 0 - 0.466] as predictors of deviation from the screen center, while video, subject, and Z-scored OSPAN were treated as random effects. There was a significant effect of Location, but no effects of N-back performance or Map performance, and these results are shown in more detail in Table 2.7 below. Importantly, however, the direction of the Map performance trend was in the expected direction (i.e., increasing deviation from center as Map performance increased), though it was not statistically significant. Figure 2.10 shows the best-fit model that included an interaction term. This was done to visualize the interaction effect to have a better understanding of what happened in the data, which made an interaction term critical.

As shown in Figure 2.10, Kansas significantly and consistently had higher deviation than Kyoto, which matches our earlier results (in Figure 2.5). Interestingly, Figure 2.10 shows that at lower levels of map performance, as map performance increased there seemed to be an increasing trend in deviation, consistent with main effect of the map task in increasing attentional breadth. Additionally, at the lowest levels of Map task performance (0, and 0.1), deviation from center seemed to slightly increase with increasing N-back performance. This is inconsistent with the main effect of the N-back task on deviation, which was to decrease it. Nevertheless, at higher levels of map performance (0.3, 0.4, 0.46) we see an apparent opposite trend, in which higher N-back performance led to decreased proportion of deviation from center, consistent with the main effect of N-back on deviation. If we consider the difference in performance in the map task and

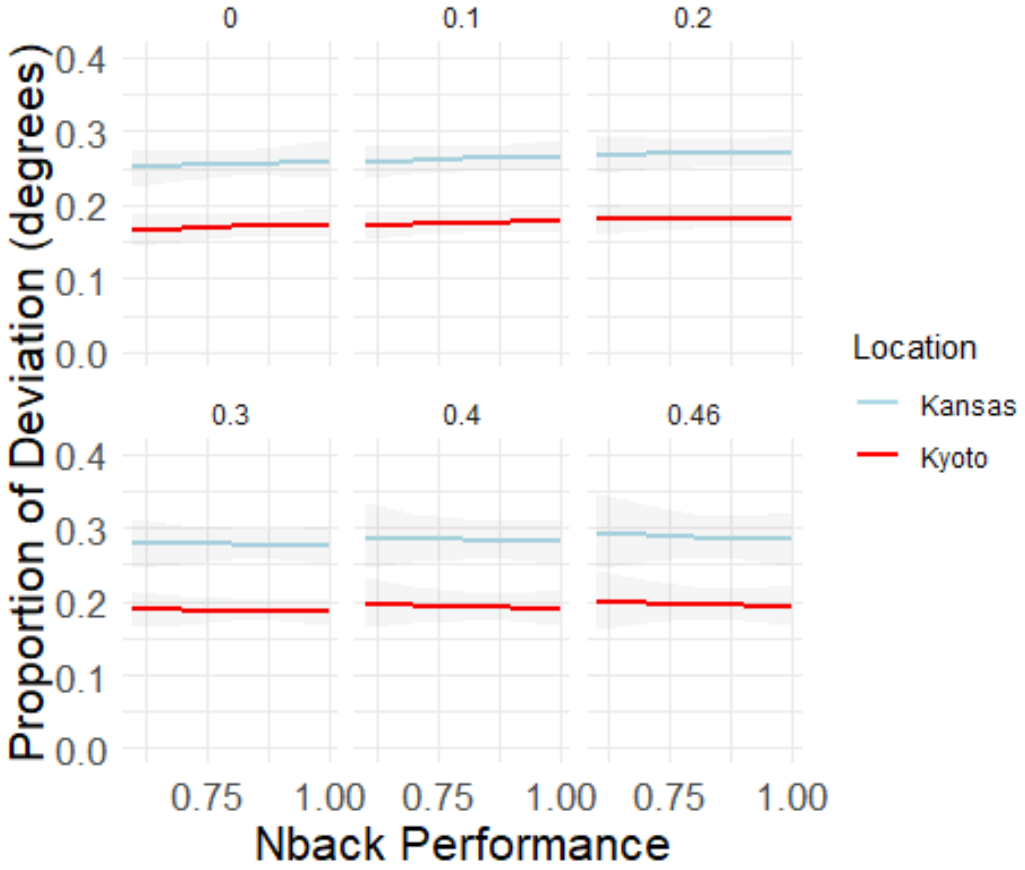
deviation trends for Kyoto and Kansas, it could be that Kyoto participants are driving the trends at higher levels of map performance, since they were performing higher overall. This could explain why we see opposite trends at these different levels.

N-back performance was neither significant nor trending towards significance in this model, which is inconsistent with our main effect of Cognitive Load on proportion of deviation and narrative-AOI fixation counts, and previous results that have shown cognitive load decreases attentional breadth (Recarte & Nunes, 2003; Reimer, 2009; Loschky, et al., 2014). These inconsistent findings could be the result of the way that N-back performance was treated. Specifically, in this exploratory analysis N-back performance was averaged over each trial for each participant (i.e., 4 data points per participant). If we revisit the original analysis (see Figure 2.5) there is evidence that attentional breadth did decrease when under the N-back compared to when not under the N-back. In this earlier analysis, we were not considering N-back performance as a predictor variable, but we did still see an impact of the overall presence of the cognitive load N-back task. Additionally, when we look at the map task performance (see Figure 2.9) we see similar evidence that performance decreased with the N-back cognitive load present. In both examples we were more globally considering the presence of the N-back task rather than the performance of each participant. Conversely, in this analysis because we have used a more simplified and basic measure of N-back (e.g., an average over each video) we could therefore have lost the chance of finding an effect of N-back performance. More specifically, the number of data points that we have in this final analysis is much fewer because we were averaging across so many trials. Another explanation could be that since Kyoto participants have lower deviation on average compared to Kansas, but their N-back performance is quite similar, the results are

negating each other and not allowing for a clear pattern to be seen. This would be especially true with the averaged N-back performance that is being used in this analysis.

Figure 2.10.

Results of Generalized Multilevel Model for Task-Trade off Analysis



Notes. Results of generalized multilevel model with the model that included the N-Back Performance X Map Performance interaction, which was the model that was used for visualization of a possible task trade-off interaction. The outcome variable was proportion of Deviation from Screen Center (degrees of visual angle). The values here represent the model predictions and error bars are 95% confidence intervals.

Table 2.6.

Task Trade-off Results

Parameter estimates Location, N-Back Performance, and Map Performance predicting Proportion of Deviation from Screen Center.

Term	Estimate	*SE*	*t*	*p*
Intercept	0.496	0.057	6.652	<.001*
Location [Kyoto]	-0.093	0.014	-6.811	<.001*
N-back Performance	0.013	0.068	0.198	.843
Map Performance	0.063	0.058	1.091	.275

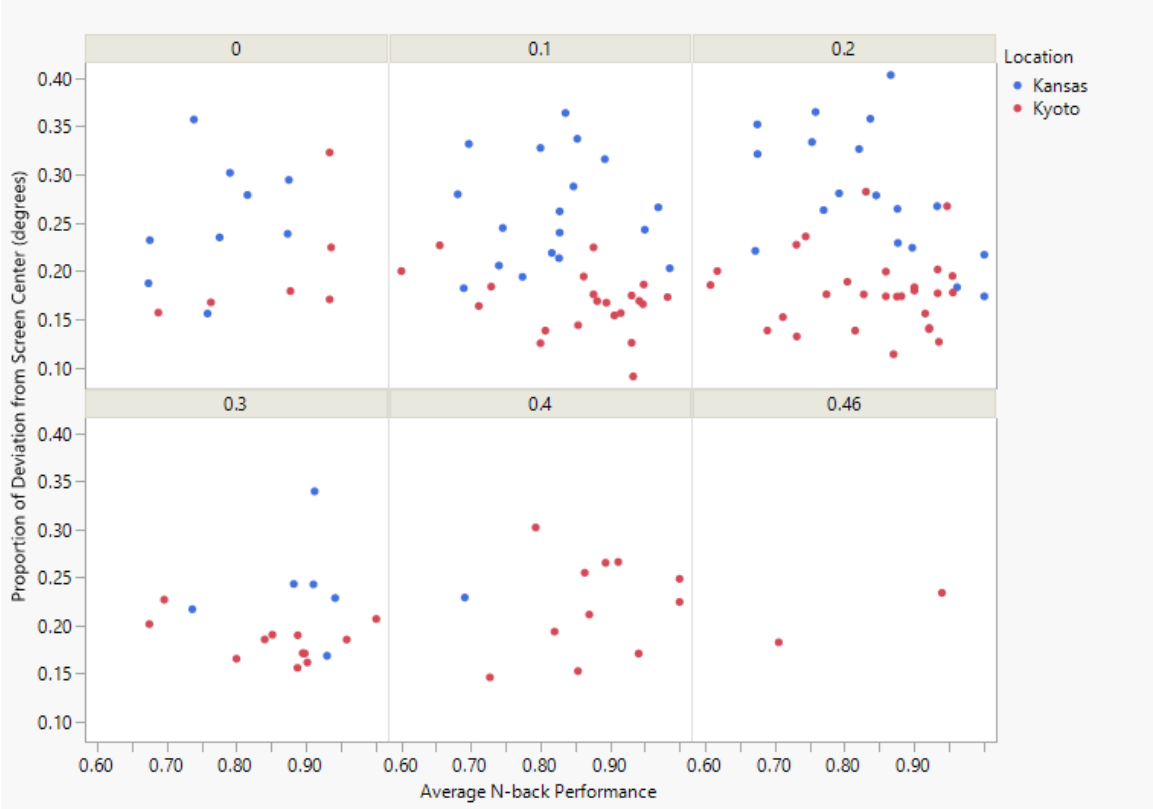
*Note: Values denoted by * are significant at the p<.05 level.*

One further possibility we investigated was based on the difference in accuracy we found between cultures in the map task. Since Kansas had significantly lower map performance on average, it is possible that there was little data for Kansas at the highest levels of map performance shown in Figure 2.10 and in the related analysis. This could mean that the multilevel model we used extrapolated beyond the raw data at those higher levels of map performance. To test this hypothesis, we plotted the raw values of deviation from the screen center for the same analyses, as shown in Figure 2.11. We see that at the highest two levels of map performance (>0.3) there was one Kansas participant but 13 Kyoto participants. This suggests that to decrease their cognitive load in this condition, which we had predicted would produce the greatest demands on executive attention, the Kansas participants gave up on the map task. This also supports the idea, given above, that our multi-level model shown in Figure 2.10 extrapolated beyond the data we had. Specifically, this helps explain why our main effects model was the best fit model, because the model was not able to estimate slopes in performance at the higher levels of map performance because of a lack of data for Kansas participants. This also seems to explain why Kansas had higher deviation overall, yet lower map scores, and they were not as impacted by the dual Map task and N-back cognitive load at these higher levels of map

performance. Specifically, as shown in Figure 2.11, the Kansas participants did not show such cognitive load effects on their deviation from center because they traded off map task performance for N-back performance, consistent with the *reserving resources hypothesis*.

Figure 2.11.

Raw Data for Task Trade-off Analysis



Notes. Depicted are the raw values for the final task trade-off analysis. N-back performance is averaged across video for participants and only those trials where N-back was present in the Map condition are included here.

Discussion

In sum, there was evidence to support pieces of our previous results. Specifically, there was a difference in eye-movements between cultures that matches our previous 3-way interaction, though Map task performance and N-Back performance were not significant

predictors. Most importantly, there was greater evidence of Kansas participants trading off performance on the Map task to maintain performance on the N-back task than for our Kyoto participants. Nevertheless, the results that support this idea were not statistically significant, but rather we have plotted the raw participant means for visual inspection of the "ground truth" of our data (see Figure 2.11). Since the data in this final analysis was averaged and reduced significantly from our original data, this could explain why did not find a significant interaction in our model; there was not enough data for the model to create reliable predictions. Nevertheless, the data visualization in Figure 2.11 is important for the overall conclusions that can be drawn from the results in Figure 2.10. Specifically, Figure 2.11 highlights that our GMLM in Figure 2.10 extrapolating a bit beyond the data, invalidating any final conclusions drawn from that analysis. Figure 2.11 also shows clear evidence consistent with the *reserving resources hypothesis*.

Globally, Figure 2.11 shows that Kyoto participants appeared to try harder in both tasks they were given, which led them to have a larger cognitive load, and thus impacted their eye-movements more than our Kansas participants. This task trade-off could have happened for several reasons, which we discuss below, but it ultimately impacted participants' breadth of attentional selection in the dual task interference paradigm.

General Discussion

Mandatory (culturally-drive) top-down effects on attention

Our data showed reliable effects of culture on attention. Our Kyoto participants' eye-movements deviated from the screen center relatively less than our Kansas participants. But our Kyoto participants spent less time focused on the central narrative elements (e.g., the main characters), which are presumably relevant to narrative comprehension, than our Kansas

participants. These results are consistent with our past research (Simonson et al., in preparation), which provide partial support for both sides of the current debate on the role the culture, particularly Western versus East-Asian cultures, on attentional selection. Specifically, four previous studies have shown no cultural differences in the degree to which people from Western and East-Asian cultures attend to central or peripheral scene content, as measured by the number of fixations to central AOIs versus the background (i.e., everything other than the central AOIs; Evans et al., 2009; Goh, Tan, & Park, 2009; Rayner et al., 2009; Masuda, Ishii, & Kimura, 2016 [for change detection pictures *having* changes]). Conversely, four studies have shown evidence that East-Asian participants made more fixations in the peripheral regions than in the central AOIs, or that the proportion of central/peripheral AOI fixations for East-Asians was lower than for Westerners (Chua, Boland, & Nisbett, 2005 [during 1,100-3,000 ms after scene onset]; Masuda, Ishii, & Kimura, 2016 [for change detection pictures *without* changes]; Rayner et al., 2007; Senzaki, Masuda & Ishii, 2014 [Exp 2, narrative generation task]). Our study did find differences in the type of information attended to by each cultural group/location, with Kansas participants making more fixations in the central narrative AOIs than the Kyoto participants. One might be tempted to interpret this as indicating that the Kyoto participants attended more broadly within the film clips. A related previous finding by Goh et al. (2009) was that East-Asian participants' eye movements covered greater distance within images than the Western participants, consistent with cultural differences in breadth of attention (e.g., East Asian participants attending more broadly to the entire scene than Westerners). However, we did not find evidence for that in our study. Instead, as noted above, our Kyoto participants showed less deviation from screen center than our Kansas participants. Thus, our study showed evidence in favor of cultural differences in attention, but not quite in the way that other studies have argued

for in the past research (Masuda & Nisbett, 2001; Nisbett & Masuda, 2003; Chua, Boland, & Nisbett, 2005; Goh, Tan, & Park, 2009). Rather, we found that when viewing film clips, participants from both cultural groups/locations primarily looked within the central portion of the screen, as shown by rather minimal deviation from screen center, but differed culturally in terms of the content that they attended to.

Importantly, this is the first investigation of cultural background, a type of mandatory top-down influence on attention that has used cinematic film as the medium for investigation (though see Kardan, et al. 2017 for everyday event videos). This is important because it shows that top-down influences can have an impact on attentional selection when viewing film, lessening the tyranny of film; something that has been difficult to find (Loschky et al., 2015; Hutson et al., 2017).

An interesting alternative explanation of our finding cultural differences in attentional selection is in terms of *culturally-specific forms* of the tyranny of film. Note that the tyranny of film has previously been defined as strong attentional synchrony among film viewers (i.e., looking at the same places at the same times), which does not allow large differences in viewers' understanding of the film to influence their attentional selection (Loschky et al., 2015; Hutson et al., 2017). But perhaps such homogenization of viewers' attention while watching film differs from culture to culture. Following this logic, it is important to note that most prior studies of cultural differences in attention used static images, which produce less attentional synchrony than dynamic images (i.e., film/video; Dorr et al., 2010). Thus, by using film clips in our study, we presumably created more attentional synchrony than those previous cross-cultural studies of attention. If so, then individual differences among viewers' attentional selection that were present in past cultural studies may have been suppressed by the bottom-up features of the film

stimuli in our study, thereby allowing stronger within-culture homogenization of attention, and producing stronger across-cultural differences in attention to be found. If so, then, according to this argument, finding cultural differences in attentional patterns while viewers watch film could be argued to be in opposition to the idea of mandatory top-down influences on attentional selection. However, a problem for such an explanation of our results is that culturally-specific forms of the tyranny of film cannot be strictly stimulus-based, given that the film stimuli in the current study were identical across both cultural groups/locations. Thus, if there are culturally-specific forms of the tyranny of film, then they must inherently involve both a stimulus-driven bottom-up component that is mapped onto a culturally-learned mandatory top-down component. For this reason, future research should investigate if the tyranny of film is culturally-specific. One way of doing so would be to determine whether there is greater attentional synchrony within each cultural group than between cultural groups when viewing the same films.

An alternative possibility is that the effects of culture we found may have been influenced, and perhaps minimized by our film stimuli. Specifically, since the current study used only Western-European film clips, and thus did not include any Japanese films, there could be a stimulus effect. For this reason, investigating the impact of film production on attentional selection in different cultures would be a highly relevant issue. For the visual narrative medium of comics, studies have shown large quantifiable differences between the *visual language* of Japanese *manga* and American comics (Cohn, 2010), and the same could presumably be true of Japanese versus Western European film. A preliminary analysis showed that our video clips did not significantly predict the proportion of deviation from screen center, but the effect was trending in a weak direction of significance ($p = .165$), which was somewhat lessened when broken down by cultural group/location (Kansas, $p = .253$; Kyoto, $p = .232$). While it is possible

that Japanese participants would be adequately exposed to Western European type film production; they may not have had as much exposure to such films as they have had to Japanese style films. This, in turn, might have caused their viewing patterns to be slightly different than normal Japanese cultural viewing patterns. This general idea is supported by the results of Ueda and Komiya (2012), who found that when Japanese participants were primed by seeing > 200 Japanese suburban and city images, and were then shown culturally neutral images (e.g., a dog in grass, or a bowl of fruits) their eye movement patterns were more expansively investigative (i.e., the holistic viewing pattern argued to be common among East-Asians), than when they were primed by > 200 American suburban and city images. Namely, exposure to one's own culturally-specific images may invoke one's culturally-specific attention patterns. This in turn suggests that all participants in the current study could have produced more centrally focused eye-movements, at least in part, due to the Western-European produced film stimuli that were chosen for this study. If so, somewhat different patterns of eye movements might be found for our Kyoto participants if shown Japanese style films instead. This also raises the question of whether Kansas participants would show a different pattern of eye movements to Japanese films.

Volitional (task-driven) attention

We replicated past research that showed volitional attention can break or attenuate the tyranny of film during film viewing (Hutson et al., 2017) and we extended that study by using multiple film clips and participants from two different cultures. Specifically, when engaged in the map task, participants willfully shifted their attention away from the main narrative elements of the film and to the background objects and locations instead. Additionally, we found evidence that such volitional attention was cognitively demanding for viewers to engage in, though direct evidence for this only came from our Kyoto participants. Nevertheless, we found that Kyoto

participants did better on both the primary and secondary tasks than Kansas participants, which suggests that the Kyoto participants were under more cognitive load. This would explain why the Kyoto participants were more impacted by the dual-task inference paradigm. Conversely, the Kansas participants did not be complete the primary and secondary tasks to the same level, but instead showed evidence of a task trade-off that reduced their cognitive load. Specifically, Kansas participants showed a greater decrement in their map task accuracy when engaged in the N-back cognitive load task than the Kyoto participants. This allowed the Kansas participants to focus their attention elsewhere (e.g., the central narrative-AOIs), and save resources for their secondary cognitive load task, for which they performed more similarly to the Kyoto participants. Evidence of this specific task trade-off by the Kansas participants provides indirect evidence consistent with the hypothesis that the Map task was cognitively demanding for them as well.

Task trade-off discussion

There are at least two potential explanations for the task trade-off differences between our two cultural groups/locations. One possibility is that it could reflect a difference between the two locations in terms of their available working memory resources. Specifically, Kyoto participants could have had a larger store of attentional resources which allowed them to perform at a higher level before seeing a decrease in their breadth of eye-movements. In terms of Figure 2.1 from the introduction, participants from Kyoto may have had a larger "reserve capacity." This in turn, would move Kyoto participants' "red line" to the right, which would increase their performance on both primary and secondary tasks compared to our Kansas participants. Nevertheless, because the Kyoto participants showed less of a task trade-off than the Kansas participants, we found that their attentional breadth was decreased for the Map task by the

secondary cognitive load task, suggesting that they exceed their "reserve capacity" during the dual-task trials.

An alternative explanation is that our Kyoto participants could simply have put more effort into the tasks we gave them than our Kansas participants. For example, if attentional resources were roughly equal between our two locations, both should have been able to perform both tasks since we have evidence that was possible. However, Kansans may have stopped trying because the Map task became too difficult to complete with ease. This explanation would be framed in terms of societal differences between Eastern and Western cultures. It is possible that for cultural reasons, the Kyoto participants may have tried harder to complete the tasks that were given to them, while Kansas participants may have been more comfortable letting their performance slip to make completing their experimental participation a bit easier (for review of important of effort in each culture see Holloway, 1998).

Future directions & conclusions

The purpose of this study was to answer two questions: 1) Is volitional attentional selection cognitively demanding for participants while watching film? and 2) Can cultural differences in mandatory attentional selection override bottom-up stimulus control in film? Our results allow us to answer "yes" to both questions. Importantly, we also can say that the demand that is caused by the volitional control task impacted our behavioral results (e.g., eye-movements and performance) between cultures differently. Our Kansas participants prioritized one task over the other when task demands were too high, while our Kyoto participants tried their best to complete both tasks equally well.

A future direction for this line of research would be to create a paradigm that eliminates the dual task nature of the current study. A paradigm utilizing neurophysiological measures (e.g.,

fMRI, EEG) could allow us to identify differences in brain activity associated with volitional versus mandatory attentional selection during film viewing. This could open doors to better understand the volitional versus mandatory distinction between these types of top-down attention.

Importantly, future research should also work to determine if the apparent "lessening" of the tyranny of film is due to a true decrease in the control of the stimulus or a difference in what the tyranny of film looks like for different cultures. While our results suggest that the tyranny of film was lessened, the alternative explanation of culturally-specific forms of the tyranny of film has not been considered a possibility until now. This is a fascinating possibility to research going forward, to identify which is the case.

Importantly, these two types of top-down attention need to be further distinguished within the attentional selection literature. Most research groups both types of top-down attentional control together. However, the current study suggests that they each having their own separate influences on attention during film viewing. By differentiating between mandatory and volitional top-down attention influences, researchers could start to understand how top-down factors truly influence eye-movements. Since mandatory effects are, by definition, more ingrained in a person, you would have to understand what is considered a mandatory effect so that you can measure and control for it, if you are not interested in that effect. For instance, since we have shown that culture influences viewing behavior patterns, this is something that could be controlled for in future studies that are uninterested in cultural effects, to help focus the results on the manipulations and measures of interest. Overall, eye-movement research on top-down effects on attention moving forward should be sure to tease apart these two distinct modes of top-down influences.

References

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In B. N. Petrov, & F. Csaki (Eds.), Proceedings of the 2nd International Symposium on Information Theory (pp. 267-281). Budapest: Akademiai Kiado.

Bach, M. (2007). The Freiburg visual acuity test- Variability unchanged by post-hoc re-analysis. *Graefe's Archive for Clinical and Experimental Ophthalmology, 245(*7), 965-971.

Bailey, H. (2012). Computer-paced versus experimenter-paced working memory span tasks: Are they equally reliable and valid? *Learning and Individual Differences, 22*, 875-81.

Baluch, F. & Itti, L. (2011). Mechanisms of top-down attention. *Trends in Neuroscience, 34*(4), 210-224.

Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye-movements during scene perception. PNAS, *102*(35), 12629-12633.

Cohn, N. (2010). Japanese visual language: The structure of manga. In Johnson-Woods, T. (Ed). Manga: An anthology of global and cultural perspectives, (pp. 187-203). New York: Continuum Books.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O, & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769-786.

Compton, E., Betley, M., Ennis, S., & Rastogi, S. (2013) 2010 Census Race and Hispanic Origin Alternative Questionnaire Experiment (No. 211).

DeAngelus, M. & Pelz, J. B. (2009). Top-down control of eye movement: Yanbu's revisited. *Visual Cognition, 17*(6-7), 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).

Fry, E. (1970). Reading Rate in 1908. *Journal of Reading, 13*(8), 593-597.

Gardony, A.L., Taylor, H.A., & Brunyé, T.T. (2016). Gardony map drawing analyzer: Software for quantitative analysis of sketch maps. *Behavior Research Methods*. *48*(1), 151 - 177.

Godijn, R., & Kramer, A. F. (2007). Antisaccade costs with static and dynamic targets. *Perception & Psychophysics, 69*(5), 802-815.

Goh, J. O., Tan, J. C., & Park, D. C. (2009). Culture Modulates Eye-Movements to Visual Novelty. *PLoS ONE, 4* (12), e8238.

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

Holloway, S. D. (1998). Concepts of Ability and Effort in Japan and the United States. *Review of Educational Research, 58*(3), 327-345.

Hutson, J. P., Smith, T. J., Magliano, J. P., & Loschky, L. C. (2017). What is the role of the film viewer? The effects of narrative comprehension and viewing tasks on gaze control in film. *Cognitive Research: Principles and Implications, 2*(24).

Jaeggi, S. M., BuschKyotoehl, M., Perrig, W.J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory, 18*(4), 394-412.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130,* 169–183.

Lahnakoski, J. M., Glerean, E. J., Jaaskelainen, I. P., Hyona, J., Hari, R., Sams, M., … Nummenmaa, L. (2014). Synchronous brain activity across individuals underlies shared psychological perspectives. *NeuroImage,* 100, 316-324.

Lenth, Russel. (2018). Emmeans: Estimated Marginal Means. aka Least-Squares Means, R package version 1.3.0. https://CRAN.R-project.org/package=emmeans

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.

Masuda, T., & Nisbett, R. E. (2001). Attending Holistically Versus Analytically: Comparing the Context Sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology, 81*(5), 922-934.

Mitchell, J. P., Macrae, N., & Gilchrest, I. D. (2002). Working Memory and the Suppression of Reflexive Saccade lengths. *Journal of Cognitive Neuroscience, 14*(1), 95-103.

Nisbett, R. E. & Masuda, T. (2003). Culture and point of view. *PNAS, 100*(19), 11163-11170.

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin, 124*(3), 372-422.

Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology-Applied, 9*(2), 119–137.

Reimer, B. (2009). Impact of cognitive task complexity on drivers' visual tunneling. *Transportation Research Record: Journal of the Transportation Research Board*, 2138, 13–19.

Ringer, R. V., Johnson, A. P., Gasper, J. G., Neider, M. B., Crowell, J., Kramer, A. F., & Loschky, L. C. (2014). Creating a New Dynamic Measure of the Useful Field of View Using Gaze-Contingent Displays. In P. Ovarfordt & D. W. Hansen (Eds.), *Proceedings of the Symposium of Eye Tracking Research and Applications,* (pp. 59-66). Safety Harbor, FL: ACM.

Ringer, R. V., Throneburg, Z., Johnson, A. P., Kramer, A.F., & Loschky, L. C. (2016). Impairing the useful field of view in natural scenes: Tunnel vision versus general interference. *Journal of Vision, 16*(2):7, 1-25.

Schwarz, G. (1978). Estimating the Dimension of a Model, *Annals of Statistics, 6,* 461-464.

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). afex: Analysis of Factorial Experiments. R package version 0.16-1. https://CRAN.R-project.org/package=afex

Simonson, T. L., et al. (2020). Volitional and Mandatory Top-down Attentional Selection during Film Viewing: The role of a goal and culture on eye-movements. [Manuscript in preparation]. Department of Psychological Sciences, Kansas State University.

Smith, T. J. & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision, 13*(8), 1-24.

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

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, *18*(6), 643.

Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working Memory Capacity and the Antisaccade Task: Individual Differences in Voluntary Saccade Control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(6), 1302-1321.

Unsworth, N., & Engle, R.W. (2007). The Nature of Individual Differences in Working Memory Capacity: Active Maintenance in Primary Memory and Controlled Search from Secondary Memory. *Psychological Review, 114*(1), 104-132.

Wickens, C.D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). Mental Workload, Stress, and Individual Differences; Cognitive and Neuroergonomic Perspectives (346-376). Pearson Education, Inc.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4.

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

Supplementary Materials Chapter 1

Introduction

Dual-task paradigms

In a study done by Azuma, Minamota, Yaoi, Osaka, & Osaka (2014) they looked at the ability of participants to perform two tasks at once using the reading span task. This task required participants to read a series of sentences and remember a random letter or word that was presented at the end of each of these sentences until the end of each trial, which was composed of a varying number of sentences. When a trial was completed, participants were instructed to recall the string of letters or words that were presented to them throughout that trial. Participants' eyes were tracked to see if the load that was presented to them had any effect on their general reading patterns (e.g., a larger the number of words or letters to be recalled meant there was a higher load for the participant). The results showed that while performing this task, viewers made more regressions to previous words, which indicates they had a harder time comprehending the sentences normally, which was thought to have been the result of an increased cognitive load. They also found that as the working memory (WM) load increased, these regressions occurred more often, suggesting a lack of executive attentional control during this task.

Method

Full video citations

1. Iñárritu, A. G. (2014). Birdman. [Film]. *Regency Enterprises*
2. Cuaron, A. (2006). Children of Men [Film] *Strike Entertainment.*
3. SoKyotorov, A. (2002). The Russian Ark [Film]. *Seville Pictures*
4. Hitchcock, A. (1948). Rope [Film]. *Universal Studios Home Entertainment*
5. Mendes, S. (2015). James Bond: Spectre [Film]. *Eon Productions.*

6. Tarkovsky, A. A. (1986). The Sacrifice [Film]. *Sandrew (Sweden Theatrical).*
7. Welles, O., & Zugsmith, A. (1958). Touch of Evil [Film]. *Universal Pictures*.

Cognitive load

The manipulation of cognitive load was a within-subjects variable. The cognitive load task that was used in this experiment is the N-back task (Jaeggi et al., 2010; Ringer et al., 2014; Ringer, Throneburg, Johnson, Kramer, & Loschky, 2016). There were two different levels of cognitive load that were used [1-back, 2-back], which was done through two separate studies. The first study was completed with the 1-back, and then this was increased to a 2-back in the second study, under the pretense that the1-back may not have been a strong enough manipulation. The cognitive load manipulation was completed during film viewing on half of the trials that participants completed, either the first four film clips or the second four (see Figure 1.1). This task presented participants with an audio stream of letters, which were presented every two seconds during the film clip. The task was to make a response (either ‘yes’ or ‘no’ on a Cedrus response box) within the two seconds of the auditory letter presentation. Participants needed to make a ‘yes’ response if the current letter was also present ‘N’ letters ago, and otherwise make a ‘no’ response, which required no response.

Results

Experiment 1: Kansas

Supplementary Materials Chapter 1 Table A.1.

Parameter Estimates Predicting Proportion of Deviation

Parameter estimates of a generalized multi-level model predicting deviation from screen center (degrees)

Term	Estimate	*SE*	*t*	*p*
Intercept	0.499	0.0145	33.89	<.001*
Condition Type [Map]	0.029	0.008	3.96	<.001*

*Note: Values denoted by * are significant at the p<.05 level. Variables are effect-coded, and all referenced to the overall average saccade amplitude (intercept)*

Table A.2.

Parameter Estimates Predicting Fixation Durations

Parameter estimates of a generalized multi-level model predicting fixation durations (ms)

Term	Estimate	*SE*	*t*	*p*
Intercept	6.084	0.034	158.512	<.001*
Condition Type [Map]	-0.057	0.048	-1.186	.236

*Note: Values denoted by * are significant at the p<.05 level. Variables are effect-coded, and all referenced to the overall average saccade amplitude (intercept).*

Table A.3.

Parameter Estimates Predicting Proportion of Saccade lengths

Parameter estimates of a generalized multi-level model predicting saccade amplitude (degrees)

Term	Estimate	*SE*	*t*	*p*
Intercept	2.018	0.045	45.085	<.001*
Condition Type [Map]	0.096	0.045	2.166	.030*

*Note: Values denoted by * are significant at the p<.05 level. Variables are effect-coded, and all referenced to the overall average saccade amplitude (intercept).*

Table A.4.

Parameter Estimates Predicting AOI Dwell Time

Parameter estimates of a generalized multi-level model predicting area of interest dwell time (ms)

Term	Estimate	*SE*	*t*	*p*
Intercept	10.176	0.28	36.74	<.001*
Condition Type [Map]	-0.156	0.06	-3.58	.009*

*Note: Values denoted by * are significant at the p<.05 level. Variables are effect-coded, and all referenced to the overall average deviation from screen center (intercept).*

Table A.5.

Parameter Estimates Predicting AOI Fixation Counts

Parameter estimates of a generalized multi-level model area of interest fixation count

Term	Estimate	*SE*	*Z*	*p*
Intercept	4.306	0.187	23.026	<.001*
Condition Type [Map]	-0.067	0.0576	-1.168	.243

*Note: Values denoted by * are significant at the p<.05 level. Variables are effect-coded, and all referenced to the overall average deviation from screen center (intercept).*

Experiment 2: Kyoto

Table A.6.

Parameter Estimates Predicting Proportion of Deviation

Parameter estimates of a generalized multi-level model predicting deviation from screen center (degrees)

Term	Estimate	*SE*	*t*	*p*
Intercept	0.439	0.013	33.359	<.001*
Condition Type [Map]	0.024	0.008	3.113	.002*

*Note: Values denoted by * are significant at the p<.05 level. Variables are effect-coded, and all referenced to the overall average deviation from screen center (intercept).*

Table A.7.

Parameter Estimates Predicting Fixation Durations

Parameter estimates of a generalized multi-level model predicting fixation durations (ms)

Term	Estimate	*SE*	*t*	*p*
Intercept	19.342	0.324	59.635	<.001*
Condition Type [Map]	1.055	0.389	2.721	.007*

*Note: Values denoted by * are significant at the p<.05 level. Variables are effect-coded, and all referenced to the overall average saccade amplitude (intercept).*

Table A.8.

Parameter Estimates Predicting Proportion of Saccade lengths

Parameter estimates of a generalized multi-level model predicting saccade amplitude (degrees)

Term	Estimate	*SE*	*t*	*p*
Intercept	0.451	0.009	52.140	<.001*
Condition Type [Map]	0.399	0.009	4.879	<.001*

*Note: Values denoted by * are significant at the p<.05 level. Variables are effect-coded, and all referenced to the overall average deviation from screen center (intercept).*

Table A.9.

Parameter Estimates Predicting AOI Dwell Time

Parameter estimates of a generalized multi-level model predicting area of interest dwell time (ms)

Term	Estimate	*SE*	*t*	*p*
Intercept	164.14	17.13	9.582	<.001*
Condition Type [Map]	-18.54	7.61	-2.434	.015*

*Note: Values denoted by * are significant at the p<.05 level. Variables are effect-coded, and all referenced to the overall average deviation from screen center (intercept).*

Table A.10.

Parameter Estimates Predicting AOI Fixation Counts

Parameter estimates of a generalized multi-level model area of interest fixation count

Term	Estimate	*SE*	*Z*	*p*
Intercept	4.28	.207	20.637	<.001*
Condition Type [Map]	-0.41	.86	-4.764	<.001*

*Note: Values denoted by * are significant at the p<.05 level. Variables are effect-coded, and all referenced to the overall average deviation from screen center (intercept).*

Experiment 2a: cultural comparison

Deviation from Screen Center. Results from a GMLM are depicted in Table A.11. We found a significant effect of condition, $\beta = 0.531$, $t = 4.132$, $p < .001$, such that those in the comic condition ($M = 0.47$, $SE = 0.01$) had less deviation compared to those in the map condition ($M = 0.49$, $SE = 0.01$); in support of the condition hypothesis. We also found support for the cultural dependence hypothesis, $\beta = 0.434$, $t = -7.887$, $p < .001$. Those from Kansas ($M = 0.52$, $SE = 0.01$) had more deviation compared to those from Kyoto ($M = 0.45$, $SE = 0.01$). This cultural difference was the same for both conditions, $\beta = .495$, $t = -0.504$, $p = .615$.

Fixation Durations. Results from a GMLM are visually depicted in Table A.12. Condition type on fixation durations showed a significant effect of condition, β = 20.454, t = -1.813, p = .007, such that those in the comic condition (M = 20.18, SE = 0.24) had shorter fixation durations compared to the map condition (M = 20.42, SE = 0.28); supporting the condition hypothesis. Consistent with the cultural dependence hypothesis, we found that those in Kansas (M = 20.74, SE = 0.23), β = 19.33, t = -4.844, $p < .001$, had longer fixation durations compared to those from Kyoto (M = 19.85, SE = 0.28); and also, with the interaction between Location and Condition, β = 22.649, t = 1.884, p = .05, such that there was less difference for Kyoto in the two locations compared to Kansas.

Saccade Amplitudes. Results from a GMLM are visually depicted in Table A.13. Condition type on saccade amplitude showed a significant effect of condition, β = -1.504, t = 2.406, p =.016, such that those in the comic condition (M = -1.6, SE = 0.03) had shorter saccade lengths compared those in the map condition (M = -1.47, SE = 0.03), in support of the condition hypothesis. There was not support for the cultural dependence hypothesis when looking at location on saccade amplitudes, β = -1.595, t = 0.21, p = .983, such that those from Kansas (M = 0.46, SE = 0.01) had similar saccade lengths compared to Kyoto (M = 0.47, SE = 0.01); there was also no interaction between Location and Condition, β = -1.519, t = 1.359, p =.174.

Area of Interest Dwell Time. Results from a GMLM are depicted in Table A.14. Condition type on the AOI dwell time resulted in a significant effect, β = 9.74, t = -4.105, p < .001, such that those in the comic condition ($M = 9.76$, $SE = 0.12$) dwelled longer in the AOIs compared those in the map condition ($M = 9.59$, $SE = 0.11$), in support of the condition hypothesis. There was support for the cultural dependence hypothesis when looking at the main effect of location, β = 9.64, t = -39.701, $p < .001$, such that those from Kansas ($M = 9.82$, SE =

0.11) dwelled longer in AOIs compared to those from Kyoto (*M =9.55, SE = 0.11*); and, for the interaction between Location and Condition, β =9.87, t = -2.691, p = .007.

Area of Interest Fixation Count. Results from a GMLM are depicted in Table A.15. Condition type on the number of fixations to an AOI resulted in a marginally significant effect, β = 3.89, Z = -1.93, p = .054, such that those in the comic condition (M = 3.96, SE = 0.10) made more fixations to AOIs compared those in the map condition (M = 3.76, SE = 0.11), in support of the condition hypothesis. The results also suggest support for the cultural dependence hypothesis when looking at location, β = 3.92, Z = -80.51, $p < .001$, such that those from Kansas (*M = 3.95, SE = 0.10*) made more fixations to AOIs compared to those from Kyoto (*M = 3.77, SE = 0.10*); and when looking at the interaction between Location and Condition, β = 3.81, Z = -110.99, p < .001.

Table A.11.

Parameter Estimates for Proportion of Deviation from Center (Cultural Comparison)

Parameter estimates of a generalized multi-level model predicting deviation from screen center (degrees)

Term	Estimate	*SE*	*t*	*p*
Intercept	0.501	0.014	36.507	<.001*
Condition Type [Map]	0.029	0.01	4.132	<.001*
Location [Kyoto]	-0.062	0.01	-7.887	<.001*
Condition [Map] X Location [Kyoto]	-0.005	0.01	-0.504	.615

*Note: Values denoted by * are significant at the p<.05 level. Variables are effect-coded, and all referenced to the overall average deviation from screen center (intercept).*

Table A.12.

Parameter Estimates for Fixation Durations (Cultural Comparison)

Parameter estimates of a generalized multi-level model predicting fixation durations (ms)

Term	Estimate	*SE*	*t*	*p*
Intercept	21.027	0.217	96.767	< .001*
Condition Type [Map]	-0.571	0.212	-2.697	.007*
Location [Kyoto]	-1.697	0.303	-5.596	<.001*
Condition [Map] X Location [Kyoto]	1.622	0.269	6.027	< .001*

*Note: Values denoted by * are significant at the p<.05 level. Variables are effect-coded, and all referenced to the overall average fixation duration (intercept).*

Table A.13.

Parameter Estimates for Proportion of Saccade lengths (Cultural Comparison)

Parameter estimates of a generalized multi-level model predicting saccade amplitude (degrees)

Term	Estimate	*SE*	*t*	*p*
Intercept	-1.596	0.04	-40.403	<.001*
Condition Type [Map]	0.0916	0.04	2.406	.016*
Location [Kyoto]	0.001	0.04	0.021	.982
Condition [Map] X Location [Kyoto]	0.078	0.06	1.359	.174

*Note: Values denoted by * are significant at the p<.05 level. Variables are effect-coded, and all referenced to the overall average saccade amplitude (intercept).*

Table A.14.

Parameter Estimates for AOI-Dwell Time (Cultural Comparison)

Parameter estimates of a generalized multi-level model predicting area of interest dwell time (ms)

Term	Estimate	*SE*	*t*	*p*
Intercept	9.895	0.107	92.877	<.001*
Condition Type [Map]	-0.159	0.039	-4.105	<.001*
Location [Kyoto]	-0.256	0.006	-39.70	<.001*
Condition [Map] X Location [Kyoto]	-0.027	0.010	-2.691	.007*

*Note: Values denoted by * are significant at the p<.05 level. Variables are effect-coded, and all referenced to the overall average AOI-dwell time (intercept).*

Table A.15.

Parameter Estimates for AOI-Fixation Count (Cultural Comparison)

Parameter estimates of a generalized multi-level model area of interest fixation count

Term	Estimate	*SE*	*t*	*p*
Intercept	3.995	0.098	40.719	<.001*
Condition Type [Map]	-0.099	0.051	-1.929	.054
Location [Kyoto]	-0.080	0.001	-80.51	<.001*
Condition [Map] X Location [Kyoto]	-0.183	0.002	-110.99	<.001*

*Note: Values denoted by * are significant at the p<.05 level. Variables are effect-coded, and all referenced to the overall average AOI-fixation count (intercept).*

Eye-movement & cognitive load analyses

Table A.16.

Parameter Estimates for Fixation Durations (Cognitive Load Included)

Parameter estimates of a full-factorial generalized multilevel model predicting fixation durations (ms).

Term	Estimate	*SE*	*t*	*p*
Intercept	20.301	0.176	115.675	<.001*
Condition Type [Comic]	-0.0119	0.171	-0.698	.485
Cognitive Load [Present]	0.220	0.015	16.033	<.001*
Data Collection Location [Kansas]	0.447	0.148	3.028	.002
Condition Type [Comic] X Cognitive Load [Present]	-0.022	0.014	-1.542	.123
Condition Type [Comic] X Location [Kansas]	0.404	0.169	2.396	.017
Cognitive Load [Present] X Location [Kansas]	0.061	0.0143	4.296	<.001*
Condition Type [Comic] X Cognitive Load [Present] X Location [Kansas]	-0.024	0.0143	-1.686	.092

*Note: Values denoted by * are significant at the p<.05 level. Variables were effect-coded, and all referenced to the overall average fixation duration (intercept).*

Table A.17.

Parameter Estimates for Proportion of Saccade lengths (Cognitive Load Included)

Parameter estimates of a full-factorial generalized multilevel model predicting proportion of saccade amplitudes (degrees of visual angle).

Term	Estimate	*SE*	*t*	*p*
Intercept	0.467	0.007	64.02	<.001*
Condition Type [Comic]	-0.0154	0.003	-4.599	<.001*
Cognitive Load [Present]	-0.004	4.5e-4	-9.187	<.001*
Data Collection Location [Kansas]	0.005	0.003	-1.394	.163
Condition Type [Comic] X Cognitive Load [Present]	1.9e-4	4.5e-4	0.435	.6637
Condition Type [Comic] X Location [Kansas]	0.005	0.003	1.376	.169
Cognitive Load [Present] X Location [Kansas]	-0.002	4.5e-4	-4.886	<.001*
Condition Type [Comic] X Cognitive Load [Present] X Location [Kansas]	0.001	4.5e-4	2.314	.207*

*Note: Values denoted by * are significant at the p<.05 level. Variables are effect-coded, and all referenced to the overall average saccade amplitude (intercept).*

Table A.18.

Estimates for Proportion of Deviation (Cognitive Load Included)

Parameter estimates of a full-factorial generalized multilevel model predicting proportion of deviation from screen center (degrees of visual angle).

Term	Estimate	*SE*	*t*	*p*
Intercept	0.483	0.013	37.379	<.001*
Cognitive Load [Present]	-0.008	3.3e-4	-24.919	<.001*
Condition Type [Comic]	-0.014	0.003	-4.999	<.001*
Data Collection Location [Kansas]	0.032	0.003	11.726	<.001*
Condition Type [Comic] X Cognitive Load [Present]	4.5e-4	3.3e-4	-1.387	.165
Condition Type [Comic] X Location [Kansas]	-0.001	0.003	-0.297	.619
Cognitive Load [Present] X Location [Kansas]	-0.002	3.3e-4	-6.939	<.001*
Condition Type [Comic] X Cognitive Load [Present] X Location [Kansas]	3.8e-4	3.3e-4	1.171	.241

*Note: Values denoted by * are significant at the p<.05 level. Variables are effect-coded, and all referenced to the overall average deviation from screen center (intercept).*

Supplementary Materials Chapter 2

Methods

Video clip details

1. Iñárritu, A. G. (2014). Birdman. [Film]. *Regency Enterprises*
2. Cuaron, A. (2006). Children of Men [Film] *Strike Entertainment.*
   a. Two clips taken from this film
3. SoKyotorov, A. (2002). The Russian Ark [Film]. *Seville Pictures*
4. Hitchcock, A. (1948). Rope [Film]. *Universal Studios Home Entertainment*
5. Mendes, S. (2015). James Bond: Spectre [Film]. *Eon Productions.*
6. Tarkovsky, A. A. (1986). The Sacrifice [Film]. *Sandrew (Sweden Theatrical).*
7. Welles, O., & Zugsmith, A. (1958). Touch of Evil [Film]. *Universal Pictures*.

Results

Data Preparation

Table B. 1.

Correlations between main outcome measures from past research (Simonson, et al., in prep) used to determine what outcome measure to focus on in current study.

	Proportion of Degrees from Screen Center	Proportion of Saccade lengths
Proportion of Degrees from Screen Center	--	--
Proportion of Saccade lengths	.2056*	--
Fixation Durations	.0152*	-.0443*

Notes. Values denoted by * are significant at the p <.001 value

Table B. 2.

Correlations between main narrative-AOI outcome measures from past research (Simonson, et al., in prep) used to determine what outcome measure to focus on in current study.

	Proportion of Fixation to Narrative-AOI	Narrative-AOI Dwell Time
Narrative-AOI Dwell Time	.9223*	--
Narrative-AOI Fixation Count	.9891*	.8794*

Notes. Values denoted by * are significant at the p <.001 value

N-Back analyses

Hypotheses

- *Null Hypothesis:* Predicts there will be no culture, condition, time, or working memory-based differences between N-back performance and no impact of N-back performance on eye-movements. This would be expected if the cognitive load was not large enough to increase the demand that participants were under, or if participants were overwhelmed because of a cognitive load that was too high.
- *Time-based Performance*: Predicts there will be a difference in N-back performance over time, either an increase (e.g., practice effects) or decrease (e.g., fatigue effects) in performance.
- *Condition-based Performance*: Predicts the condition participants are in will influence N-back performance. Specifically, since the map task is hypothesized to be more demanding than the comic task, it would be expected that people performing the map task would have fewer working memory resources to devote to the N-back task (Mitchell, Macrae, Gilchrist, 2002; Unsworth, Schrock, & Engle, 2004).
- *Culture-based Performance:* Predicts there is a difference in N-back performance related to the participants'' cultural background. Specifically, Kyoto participants would be performing better, which would explain why they were exhibiting more cognitive load in the 3-way interaction.
- *Working Memory Dependence Performance*: Predicts that there is a role for individual E-WM capacity in N-back performance ability. Specifically, we would expect those with higher E-WM capacity to perform more accurately across the board (Mitchell, Macrae, Gilchrist, 2002; Unsworth, Schrock, & Engle, 2004).

Performance

This model was run using JMP software and no model comparison techniques were used. We were only interested in the general trends of the data, so the main effects were included without any interaction effects. Video and subject were entered as random effects (RE) at their intercept, similarly to other analyses using this data**.**

OSPAN

Several models were compared for the one that fit the data the best. The first two models included fixed effects of Condition, Location, and Z-scored OSPAN, and included RE of Video and Subject at the intercept, while model 1 used a square root link function and model 2 used a log link function. The next two models included full-factorial fixed effects of Condition, Location, and Z-scored OSPAN, and the RE of Video and Subject at the intercept, while model 3 used a square root link function and model 4 used a log link function. Both full-factorial models failed to converge and will not be discussed any further. AIC and BIC values can be seen below in Table B.3. The best fit model for this data was model 1, which utilized a generalized multilevel model with a gamma family and square root link function.

Table B.3.

Model Comparisons for Models

Model	AIC	BIC
Model 1	-762.1	-735.9
Model 2	-756.2	-730.1

Note: Model 1 had a square root link function and Model 2 had a log link function.

Using generalized multilevel modeling, Condition [Map =1, Comic = 0], Location [Kyoto = 1, Kansas = 0], and Z-score OSPAN [range: -2.14 – 2.15] were used to predict N-back

Performance [range: 0.5 - 1]. Here N-back Performance is being operationalized as the overall percentage of accuracy on each trial with N-back that the participants; this means that each participant will have four points in the data. For a discussion on how the model was chosen, see supplementary materials.

Results are visually depicted in Table B.4 below and show a significant effect of Location, such that participants from Kansas (M = 0.89, SE = 0.01) performed significantly worse than those participants from Kyoto (M = 0.92, SE = 0.01); in support of the *culture-based performance hypothesis.* Results also show a significant effect of Z-Scored OSPAN, such that as participants OSPAN score increased their performance on the N-back task increased also. This is in support of the *working-memory dependent hypothesis* that predicted working-memory capacity would play a role in the performance ability of participants. There was no effect of the condition that participants were in when predicting N-back Performance, which is not in support of the *condition-based performance hypothesis.*

Table B.4.

Parameter Estimates Predicting N-Back Performance

Parameter estimates Condition, Location, and Z-score OSPAN predicting N-Back Performance.

Term	Estimate	*SE*	*t*	*p*
Intercept	0.903	0.014	62.315	<.001*
Condition [Map]	-0.012	0.019	-0.645	.519
Location [Kyoto]	0.029	0.005	6.016	<.001*
Z-Score OSPAN [WM]	0.012	0.003	3.895	<.001*

*Note: Values denoted by * are significant at the p<.05 level.*

Time

Two models were compared for the best fit of the data. Both models were set with a binomial family distribution and included video and subject as RE at their intercept. The difference between the two models was that the fixed effects for model 1 were the main effects of Location, Condition, and Time, while model 2 had the full-factorial effects of these variables. AIC and BIC values can be seen below in Table B.5. The best fit model for this data was model 1.

Table B.5.

Model Comparisons for Models

Model	AIC	BIC
Model 1	20,218	20,267
Model 2	20,225	20,305

Note: Model 1 included main effects only and Model 2 was a full-factorial

Map Analyses

Data preparation: master map creation details

Master map creation. To score the maps for accuracy there must be a template that each participant map can be scored against. To do this a *master map* was created for each film, that was neutral to differences in cultural location. This process utilized a group of individuals who were familiar with the videos and ensured that they had access to the videos during the entire process. These individuals were instructed to draw a map of the video with as many objects and locations accurately mapped out as possible. They were encouraged to watch each of the film clips repeatedly to create as accurate of a map as possible. This step was done in both the Kansas

and Kyto populations. After this the first author of the paper used these maps and created a single map for each film that was based on these created maps. Similarly, this researcher watched the videos to ensure the maps were as detailed and accurate as possible. This researcher oversaw creation of both sets of master maps, to ensure that similar procedures were used across the maps and locations. After this step was complete a single map for each film clip combined all features of both maps and created a *Grand Master Map.*

Reliability. When scoring each of the participant maps, to ensure that each map is consistently scored, undergraduate research assistants were trained by an experimenter on how to use the Gardony Map Drawing Analyzer (GMDA) software. All coders practiced the same set of participant maps and practiced identifying and discussing any discrepancies that arose in the presence of a researcher. Upon completion of practice rounds, coders began scoring their assigned maps. In Kansas, coders worked with all maps between 2 coders. In Kyoto, coders worked with a partner and only completed half of the maps, or 4 videos worth. Coders went through and completed half of the maps for a given film and would discuss any discrepancies they ran into and resolve them. After resolving discrepancies, they completed the second half of the maps for that film. Scores for agreement can be found in Table B.6 below.

Table B.6.

Coding Agreements for Each Video, Split by Locations

	Round 1a	Round 1b	Round 1c	Round 2a	Round 2b	Round 2c
			Kansas Coders			
Children of Men: Building	95.80%	95.80%	100%	67.10%	97.60%	100%
Children of Men: Town	48.50%	100%	100%	60%	96.70%	100%

Rope	---	---	---	75%	95.50%	100%
The Russian Ark	70.60%	97.30%	100%	47.20%	96.40%	100%
Sacrifice	75.70%	99%	100%	97.40%	98.20%	100%
James Bond: Spectre	66.40%	98.30%	100%	80.40%	100%	100%
Birdman	66.10%	99%	100%	67.70%	100%	100%
Touch of Evil	60%	98%	100%	52.90%	100%	100%
			Kyoto Coders			
Children of Men: Building	56.32%	93.68%	100%	78.18%	98.21%	100%
Children of Men: Town	61.1%	92.02%	100%	81.68%	96.58%	100%
Rope	---	---	---	62.8%	96.24%	100%
The Russian Ark	61.96%	98.76%	100%	80.42%	92.38%	100%
Sacrifice	89.72%	100%	100%	100%	---	---
James Bond: Spectre	87.41%	98.45%	100%	86.34%	100%	
Birdman	95.45%	100%	--	97.2%	100%	--
Touch of Evil	26.92%	95.30%	100%	62.99%	96.75%	100%

Hypotheses

- *Null Hypothesis* predicts there will be no difference in map accuracy regardless of the amount of cognitive load participants were under when completing the map task.

- *Culture-based performance Hypothesis* predicts that there will be differences in performance that are explained by cultural background. Specifically, Kyoto participants are performing better compared to Kansas because of typical eye-movement results that suggest Eastern populations tend to view a scene more holistically than Western populations (Masuda & Nisbett, 2001; Nisbett & Masuda, 2003; Chua, Boland, & Nisbett, 2005) which is in-line with the result that were found in the above results (see Figure 2.5).
- *Cognitive Load Hypothesis* predicts there will be an overall higher accuracy in the map task condition when participants complete the trials that have no cognitive load.
- *E-WM Capacity Hypothesis* predicts there will be a difference in accuracy, but that this difference in accuracy will depend on the level of E-WM capacity that individuals have (Mitchell, Macrae, Gilchrist, 2002; Unsworth, Schrock, & Engle, 2004).
- *Eye-Movement influence Hypothesis* predicts that eye-movements will predict how well participants are performing the map task. Specifically, that more spread in eye-movements (e.g., increased deviation) will be required for higher map performance.

Data preparation

The output that was obtained from these map scoring procedures is referred to as the SQRT (canonical accuracy) which gives measures for both the number of locations marked and the spatial configuration of the landmarks included on the maps. This outputted report is in respect to the template Grand Master Map that each participant's map was compared to.

Performance

Two different models were compared for best fit with the data. Both models included the full-factorial main effects of Cognitive Load [Presence or Absence] and Location [Kyoto or Kansas], as well as subject and video as RE at their intercept. The difference between these

models was the link function that was used, either a gamma distribution with a square root link function or a gamma with log link function. The AIC and BIC values for both models are shown in Table B.7 below and you will see that model 2 was the best fit for the data.

Table B.7.

Model Comparisons for Models

Model	AIC	BIC
Model 1	-522.52	-495.19
Model 2	-526.31	-498.98

OSPAN

There were two model comparisons made for best model fit. Both models were gamma distributions with square root link functions and included subject and video as RE at their intercepts. Where the two models differed was in the fixed effects included in the model. Model 1 had main effects of Cognitive Load [Absent or Present], Location [Kyoto or Kansas], and Z-scored OSPAN, while model 2 had a full-factorial of those same variables. The AIC and BIC values can be found in Table B.8 below and shows that Model 1 was a better fit for the data.

Table B.8.

Model Comparisons for Models

Model	AIC	BIC
Model 1	-583.03	-558
Model 2	-579.37	-540.04

Generalized multilevel model with a gamma distribution and a square root link function was the best fit for the data, discussion of model comparison can be found in the supplementary materials. Cognitive Load [Absent = 1; Present = 0], Location [Kyoto = 1; Kansas = 0] and the Z-scored OSPAN scores [Range: -1.84 – 2.02] were entered as fixed effects, while participant and video were entered as RE at their intercept. The outcome measure in this model is map performance.

Results are depicted in Table B.9 below and show a significant effect of Location, participants from Kyoto (M = 0.46, SE = 0.03) had higher map performance compared to participants from Kansas (M = 0.39, SE = 0.03); which is in line with the *culture-based performance hypothesis*. There was not a significant effect of OSPAN scores or cognitive load presence/absence when predicting map performance. This is not in support of the *E-WM Capacity Hypothesis* which predicted a difference in accuracy that depends on the level of E-WM capacity that the individual has or the *cognitive load hypothesis.*

Table B.9.

Parameter Estimates for Map Performance

Parameter estimates Cognitive Load, Location, and Z-score OSPAN predicting Map Performance.

Term	Estimate	*SE*	t	*p*
Intercept	0.375	0.032	11.833	<.001*
Cognitive Load [Absent]	0.022	0.014	1.59	.11
Location [Kyoto]	0.070	0.012	5.67	<.001*
Z-score OSPAN [WM]	0.001	0.007	0.137	.89

*Note: Values denoted by * are significant at the p<.05 level.*

Discussion

Overall, there was support for the culture-based performance hypothesis that predicted Kyoto would have higher performance compared to Kansas. It is possible that this result was the biproduct of Kyoto participants naturally looking at the background more, making this easier for them to do better on the map task. While it is possible that this task was easier for Kyoto, they still showed the larger drop in deviation when in the dual task interference paradigm, suggesting that they were still being cognitively challenged during this map task.

There was also evidence in support of the *Cognitive Load Hypothesis,* which predicted that when under a cognitive load there would be a decrease in map performance. This is in line with past research that suggests participants do not look around as much and potentially narrow their gaze more when under an increased cognitive load (Recarte & Nunes, 2003; Reimer, 2009; Loschky, et al., 2014). Additionally, a significant Cognitive Load X Location interaction suggests that we are still in-line with that initial 3-way interaction result (see Figure 5) where Kyoto had a larger difference between their cognitive load present and absent trials during the

map task condition. Here we are also mapping these results onto the more precise map task performance measure as well.

Shockingly, there was no evidence that WM played a role in map performance. The *E-WM Capacity Hypothesis* predicted that participants who have a higher WM would have an easier time completing these tasks, resulting in higher performance overall. This was not what we found. Instead, it was shown that there was no difference between high and low WM individuals. Interestingly, WM was a significant predictor of N-back performance. This difference could be explained by the N-back measure being more directly related to WM (Mitchell, Macrae, Gilchrist, 2002; Unsworth, Schrock, & Engle, 2004), while the map task is maybe not tapping into the same measure that N-back is. Since the map task is a difficult volitional task, we are now tapping into something different that is not related to WM capacity, but rather the ability to control your eye-movements. Though this still does not completely explain the difference because this effect has been shown to be related to WM as well (Kane, Beckley, Conway, & Engle, 2001; Unsworth, Schrock, & Engle, 2004; Unsworth & Engle, 2007).

Finally, there was only trending results for the E*ye-movement Influence hypothesis*. The predicted effect was that as deviation from screen center increased the map performance scores would also increase, the thought being that to do well on the map task they should be looking around the screen more. These results were not significant, but they were trending in the expected direction, suggesting that eye-movements were slightly predictive of map performance.

Exploratory Analysis: Working Memory as Predictor

An additional analysis was done looking at the role that working memory plays in the task trade-off analysis that was done (see Figure 2.10 and Figure 2.11). In this analysis it was of

interest to see if some of the performance results could be explained by the participants individual working memory ability. Upon running this analysis to see if working memory (e.g., OSPAN scores) there was no significant, or trending, effect that suggested working memory was explaining any of the variance. Critically, our measure of working memory was completed in English, which may have put Kyoto participants at a disadvantage, even though they were all familiar with the language as a requirement for their acceptance into university. So this could be a reason why we did not find any significant results.