THROUGH THE EYES OF THE CAMERA: UNDERSTANDING SPATIAL RELATIONS AND PERSPECTIVE TAKING IN FILM

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Ayse Candan Simsek

August 2017

THROUGH THE EYES OF THE CAMERA: UNDERSTANDING SPATIAL RELATIONS AND PERSPECTIVE TAKING IN FILM

Ayse Candan Simsek, Ph. D.
Cornell University 2017

The study of edited moving images has started to attract more interest among researchers in recent years due to their complex yet highly constructed nature, especially with respect to spatiotemporal continuity. Movies are unique visual stimuli that offer an enjoyable and seamless experience in the face of an objectively detached and segmented structure.

Continuity editing rules are at the core of Hollywood cinema and those aim mainly at maintaining spatial continuity across shots. This dissertation provides further understanding to the perceptual mechanisms used to make accurate and fast integration of spatial information provided in separate movie shots into a coherent spatial representation. Those, in most cases, represent more than one agent's viewpoint. In the scope of this dissertation, four main lines of experiments are carried out to examine how editing conventions affect viewers' judgments for spatial relations, especially involving the position of actors in a movie scene.

The results indicated that the employment of the 180-degree rule facilitates viewers' judgments for actor positions in movie scenes and leads to faster decisions. In addition, establishing shots, which are wide-angle shots positioned at the beginning

of scenes, have a complimentary but important role in keeping those relations current. Results also showed that congruent agent cues (gaze and body direction) lead to more accurate and faster judgments with respect to an upcoming position of an agent and viewers put more emphasis on body direction.

Overall, the discussed experiments support the view that continuity editing rules in movies make use of people's perceptual tendencies. The strategic usage of camera angles offers better and faster solutions to complex visual information. These rules facilitate spatial transformations across shots and alleviate cognitive resources dedicated to maintaining a coherent spatial map, which is otherwise effortful. Therefore, the viewer can allocate the much-needed resources to follow the narrative more efficiently.

BIOGRAPHICAL SKETCH

Ayse Candan Simsek was born in Ankara, Turkey. She attended Notre Dame de Sion High School, from where she graduated in 2001. She has earned her Bachelor of Arts degree in Psychology at Bogazici University in 2006. Her interest in experimental research has begun in her early college years and her fascination with psychological mechanisms led her to pursue further education in this area. She has completed her Masters of Arts degree in Developmental Psychology at Koc University in 2009. She has worked on language development with Aylin Kuntay and her M.A. thesis was on the cross-cultural variation in children's acquisition of early grammatical structure. Her years in Koc University helped motivate her decision to pursue an academic career. In 2010, she started her Ph.D. at Cornell University in Psychology, concentrating on Perception, Cognition and Development. She has worked with her advisor James E. Cutting on perception of film with a focus on viewers' understanding of spatial relations and viewpoint in movies. During her years at Cornell University, she has also worked on event segmentation and memory for briefly viewed dynamic scenes. Her research interests are visual short-term memory, spatial perception, perspective taking and viewpoint dependency in complex dynamic scenes. She is currently focusing on frame of reference and integration of spatial details in visual narratives.

To my husband, my mother and father and my grandmother. Without them, none of this would be possible.

ACKNOWLEDGMENTS

First, I would like to express my deepest and sincerest gratitude for my advisor James Cutting, whose guidance and wisdom helped prepare me for my future years in the academia. I'm very grateful and thankful to him for his direction, mentorship and indispensable support during my years at Cornell. I feel very fortunate to have him as my advisor.

I also would like to thank my committee members for their crucial contributions to my research. I'd like to give many thanks to David Field for his invaluable discussions and always-constructive feedback, which provided me with a new appreciation for asking the right questions. I am thankful to Barbara Finlay and David Pizarro for sharing their unique viewpoints and support through my graduate journey.

I also would like to express my appreciation to all my professors who helped shape my future and contributed to my education greatly. I'd like to express my gratitude to Khena Swallow for her guidance and invaluable opinions on my research. It has been a pleasure to work with her. I am also very grateful to Harry Segal, for his kindness, support, and friendship. I'd like to thank my former advisor Aylin Kuntay for paving the way for my success and my former professors Sami Gulgoz and Ali Tekcan for their help and guidance.

I am very thankful for the support of the Psychology Department staff, especially Pamela Cunningham, for their invaluable assistance and help over the years.

I'd like to express my dearest thanks to my lab friends and colleagues who have helped me with various facets of my graduate years. I'd like to give many thanks to Kedar Vilankar for his indispensable help with my research and his support and encouragement. I'd like to thank Catalina Iricinschi for always being the sound of wisdom; her help will always be appreciated. I'm also very thankful to Jordan DeLong for his friendship, support and invaluable help with my research and to Kate Brunick, who shared my concerns and lifted my morale for many years.

I'd like to give my deepest and sincerest thanks to my friends Kacie and Steve Armstrong, Erin Isbilen, Melissa Elston who contributed greatly to this dissertation. I appreciate their friendship, support and generous help over the years of my graduate journey.

I also want to thank my research assistants, Clara Seo Lee, Coral Keller, Maamie-Asamoah Mensah and Cynthia Vella, for their important contributions to this dissertation.

Last but not least, I'd like to give huge thanks to my family. My parents, my sister, my dearest grandmother and my husband were always alongside me whenever I needed them through all my years at Cornell. I owe everything to their patience, love, kindness and ever-growing support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

GENERAL INTRODUCTION

> A single concept has dominated reflection on filmic space – position. … each image is attributed to an invisible observer incarnated in the camera; this observer is at once narrator and spectator.
>
> −David Bordwell (1985, p. 99)

Studying edited moving images to further understand the perception of complex dynamic visual scenes is still new. This line of research has gained momentum recently through the adoption of a framework known as the *Cognitive Film Theory* (Bordwell & Thompson, 2003; Levin & Baker, in press; Smith, 2012; Smith, Levin & Cutting, 2012).

The perception of space in movies has recently attracted more interest among researchers due to its central role in narrative continuity (Baker & Levin, 2015; Baker, Levin, & Saylor, 2016; Cumming, Greenberg, & Kelly, 2017; Hirose, Kennedy & Tatler, 2010; Huff & Schwan, 2012; Ildirar & Schwan, 2014; Levin & Wang, 2009; Schwan & Ildirar, 2010). Some of these studies have focused on object-related information across viewpoint changes in movies while others studied the role of continuity editing rules and order of action on attention. Research so far, left many unanswered questions about how viewers form coherent spatial maps of movie scenes and keep track of the relative positions of objects and agents across different camera angles.

Movies use various editing techniques and provide a rich medium to study spatial relations in complex dynamic scenes. The series of experiments discussed in the following chapters of this dissertation focus on viewers' perception of agent-object relations in movies. The main motivation is to understand which cues are essential for viewers to form an efficient spatial integration of information presented from different but carefully connected viewing angles.

The frame of reference used to encode visual scenes has interested researchers for a long time. How people represent the position of objects in space, with respect to an egocentric reference or with respect to the environment, has been conceptualized as "viewer-centered" versus "environment-centered" coding (Jiang & Swallow, 2013; Tversky & Hard, 2009). An egocentric, viewer-dependent coding of space has been considered as more cognitively efficient when an agent moves around the environment and updates his spatial map as his perspective changes with position (Jiang, Swallow & Capistrano, 2013; Shelton & McNamara, 2004; Simons & Wang, 1998; Wang & Spelke, 2000). Early research on viewpoint-dependency mostly used static visual scenes and focused on object recognition. Researchers found that people mainly relied on an egocentric viewpoint to locate objects in visual scenes (Diwadkar & McNamara, 1997; Johnston & Hayes, 2000; Lawson & Humphreys, 1996; Shelton & McNamara, 2004).

There are mixed views about how much viewpoint is central to remembering an instance from a dynamic scene. Researchers suggested that *spatial updating,* which refers to updating the relative location of objects across viewpoint changes (Wang et al. 2006), is also performed when the viewer is stationary and is exposed to viewpoint changes passively on a computer screen (Huff & Schwan 2012; Meyerhoff, Huff, Papenmeier, Jahn, & Schwan, 2011). This line of research supported a *spatial alignment* hypothesis for memory in dynamic scenes, where performance declined with increasing deviance from the original viewpoint angle (Huff, Jahn, & Schwan, 2009; Garsoffky, Schwan & Hesse, 2002). In contrast, scholars who studied event segmentation supported viewpoint-independent conclusions in visual narratives where the content is abstracted and spatial relations are used only if they are meaningful to the story (Magliano, Miller & Zwaan, 2001). More recent studies showed that semantic content of a dynamic event influenced how much viewers depended on viewpoint. Researchers showed that meaningful sematic content led to viewpoint-independent memory for a visual scene through abstraction of information (Garsoffky, Huff & Schwan, 2007; Huff, Schwan & Garsoffky, 2011).

One can conceptualize changes in viewpoint as similar to event breakpoints, which require updating the current representation of an event due to changes in the situation. According to event perception literature, event boundaries are expected to occur at situational discontinuities such as when a salient perceptual feature (motion, color, or shape) or a conceptual feature (goals of the actors) has changed (Zacks et al. 2007; Zacks, Kurby, Eisenberg & Haroutunian, 2011; Zacks, Speer, Swallow, & Maley, 2010). Consistent with this, Hard, Recchia and Tversky (2011) found that viewers spend more time looking at breakpoints in narratives. These boundaries were inspected longer suggesting a link between segmentation and attentional resources. Zacks et al. (2011) also found an increase in brain activity at event boundaries, which coincided with moments where subjects had difficulty predicting the next activity in an event sequence. The event segmentation literature has provided mixed results on the relative importance of spatial changes on event boundaries (Magliano, Miller & Zwaan, 2001; Zacks, Speer, & Reynolds, 2009). Some studies have found that spatial changes did not affect boundary decisions when they were not crucial to the comprehension of the narrative story and were only monitored when people were motivated to monitor them (Magliano et al. 2001; Zwaan & Radvansky, 1998; Zwaan et al., 1995). So, it is not clear how much effort and attention is dedicated to keeping track of spatial details in a visual narrative and how much perspective changes across shots affect processing time when one needs to predict what is most likely to happen next.

For dynamic scenes, studies showed that people found it difficult to align different viewpoints across views and alignment accuracy and speed depended on angular distance (Garsoffky, Huff, & Schwan, 2007). Early studies with simple dynamic scenes supported a spatial alignment process, where the viewer prefers and responds faster to a new perspective when it matches the older one in closer angular distance (Mou, McNamara, Valiquette, & Rump, 2004). In movies, researchers argued that the editing conventions (especially the screen direction rule) help with this spatial alignment process. Huff and Schwan (2012) used simple dynamic scenes in which a car was moving in a certain direction in one shot and appeared in a new position when the viewpoint changed across a cut. They showed that people use what the authors called a *heuristic spatial updating*. This uses the remapping of a previous shot to the next through a heuristic, which uses screen direction instead of the effortful process of spatial alignment (Huff & Schwan, 2012).

As this may be the case for simple transformations, how people combine and make sense of information presented from the viewpoint of multiple agents in dynamic scenes is not well known. Research on written narratives, which employed multiple-viewpoint descriptions of visual scenes, found that people are flexible in their descriptions of a visual scene and can switch back and forth between different viewpoints. People mostly adopted an allocentric viewpoint when multiple actors' viewpoints are described verbally (Franklin, Tversky & Coon, 1992). In relation, some researchers have argued that spatial maps do not necessarily have an inherent perspective. Those argued that visual scenes can be coded and described from more than one viewpoint (Cavallo, Capozzi, Tversky & Becchio, 2016; Franklin, Tversky & Coon 1992; Tatler & Land, 2011; Tversky & Hard, 2009) and spatial content can be abstracted from specific instances (Allen, Siegel, & Rosinski, 1978).

Visual events are highly complex. Especially, movies involve multiple viewing angles where more than one actor's viewpoint is depicted. A cinematic narrative presents a special case where the viewer is a stationary and passive observer in the face of frequent changes in camera angles. An alternate space is portrayed to the viewer in successive shots. According to Bordwell (1985), this space " ... built up from editing is then attributed to an idealized invisible witness, the occupant of an absolute position, Pudovkin's observer ideally mobile in space and time" (p. 99). This suggests that, in movies, the viewer involvement is encouraged through adopting the perspective of the camera but the viewer nonetheless remains a silent third-party observer. Bordwell also described the process of perspective taking through the camera such that "... traditional film theory ... creates a perspectival eye for cinema, one we call the invisible observer" (1985, p. 9). Because frequent perspective changes are at the core of films, understanding how those are juxtaposed to facilitate the transitions between shots and increase the involvement of the viewer into the narrative is essential.

When viewers are watching a movie, they are not exposed to each step of the story but they can still easily form a coherent narrative based on assumption of continuity. One of the first lines of evidence demonstrating how the successive shots can be subject to inference and perceptual illusions came from studies that used the *Russian montage*. In 1920s, filmmakers Kuleshov and Pudovkin inserted different themed shots before and after a shot in which an actor was staring off screen with a neutral expression. People attributed different emotional content to the actor's expression depending on the content in the surrounding shots (bowl of soup, lounging woman, dead child). This has come to known as the "Kuleshov effect" (Levin & Simons, 2000; Levin & Baker, in press). This effect shows the important role of context in how viewers interpret a combination of shots. In relation, viewers' attention to narrative over visual details is documented in various experiments where people failed to notice changes in successive movie shots, even for information that was at the center of attention (Angelone, Levin, & Simons, 2003; Levin & Simons, 1997; Messaris 1994). Nonetheless, in those cases, the spatial layout of the scene and the positioning of the characters mostly remained the same. This suggests that spatial continuity is an essential part of ensuring narrative continuity in movies.

Levin and Simons (2000) interpreted the results from change detection studies as: "… in film, assumption of continuity is an extrapolation based on very minimal information” (p. 361). This suggests that consistency of only the essential elements might be enough to preserve continuity across shots. If the narrative coherence is preserved, it appears to have priority over the perceptual inconsistencies between shots. One can even argue that disregarding the changes in the non-essential details may be a strategy one uses to allocate cognitive resources efficiently. This is also in line with a constructive approach to memory where people "fill in the blanks" of their experience based on existing previous knowledge, their schemas of a situation (Chun & Jiang, 1998; Schacter, Norman, & Koutstaal, 1998).

“The real world is spatially and temporally continuous: film is not” (Cutting, 2005, p. 9). So, how do we achieve spatial continuity across shots, which provide different visual angles? In movies, the changes in camera angles do not occur randomly but follow a calculated pattern (Bordwell & Thompson, 2003; Magliano, Dijkstra, & Zwaan, 1996; Schwan & Ildirar, 2010). Contemporary American cinema employ conventions known as the *continuity editing* rules, which are geared toward maintaining spatiotemporal continuity (Bordwell & Thompson 2003; Smith, Levin & Cutting, 2012). This approach has also been called an *invisible style* or *Hollywood style*, which focuses on minimizing the awareness of cuts to decrease the artificiality of shot sequences (Smith & Henderson, 2008). One of the main motivations behind continuity editing rules is not only to decrease the awareness of cuts but more importantly to increase narrative involvement (Magliano & Zacks, 2011; Smith & Henderson 2008; Smith, 2012). This style includes editing techniques such as 180-degree rule, shot-reverse shots, and gaze matches to promote spatial continuity across shots (Bordwell, 1985; Chandler, 2009; Magliano & Zacks, 2011; Smith, Levin &

Cutting, 2012) (please refer to the *Terms* section in Chapter 2 for a detailed description of the editing techniques that were discussed in this dissertation).

Directional continuity, which refers to the congruency of screen direction across movie shots, is mainly achieved by the usage of the *180-degree rule*. Previous studies have differed in focus, methodology and conclusions with respect to how viewers make sense of directional violations. One of the early studies that used slide presentations of still images, showed that people's memory for shots was better when the axis of action was not crossed (Kraft, Cantor & Gottdiener, 1991). Studies also examined eye movements across editing violations and showed that the frequency of eye movements increased in the first 200-400 ms after a cut, which violated the directional continuity (d'Ydewalle & Vanderbeeken, 1990). Germeys and d'Ydewalle (2007) used more controlled stimuli with a uniform background in a follow-up study. Contrary to the first study, the authors concluded that directional violations do not necessarily create confusion but eye movements show a reorientation to the newly informative parts of the screen; mainly to the character who speaks. In more recent study, Baker and Levin (2015) also examined the role of spatial relations in movie sequences to understand how they relate to change detection. They found that the frequency of change detection for object properties increased when those coincided with a directional violation. In relation, Huff and Schwan (2012) supported an efficiency model for the usage of directional continuity in dynamic scenes. Using simple animations of moving objects, they suggested that people use screen direction rule as a *heuristic* to spatial updating across angle changes. They suggested that people use respective screen direction of objects to judge motion direction instead of a more effortful updating based on spatial alignment.

In addition to the axis of action, Bordwell and Thompson (2003) suggested that establishing shots play a crucial role in defining agent-object relations in movie scenes. It was suggested that spatial information in the subsequent shots are integrated into a spatial map that approximates the one given in the establishing shot (Garsoffky, Schwan & Hesse, 2002). In that regard, establishing shots appear to be shortcuts directors use to help viewers allocate cognitive resources more efficiently. In their study, Kraft et al. (1991) showed that people relied on establishing shots to describe the overall spatial relations in movies and those were given more weight the 180-degree line was crossed.

Bordwell (2002) proposed that revisions to film techniques over the course of the last 50 years created an *intensified continuity*, which uses traditional continuity editing rules more heavily. This showed a trend for more rapid cutting, shorter average shot durations and more close-ups in conversations. Similar trends have been shown in recent studies with respect to changes in pace, motion, luminance and shot scale (Cutting, DeLong & Nothelfer, 2010; Cutting, Brunick, DeLong, Iricinschi & Candan, 2011; Cutting & Candan, 2015). There has also been debate over what it means for an editing style to be continuous. Smith (2012) suggested that, “continuity is not the style but the intended outcome of the style” and referred to continuity as “continuity of the viewer cognition” (p. 4). This *attentional theory of cinematic continuity* suggests that the viewer is an active participant in the viewing process, not just a passive receiver. Continuity is therefore achieved by synchronizing the viewer’s attention to the most relevant parts of the screen through perceptual cues such as motion, gaze, sound, gesture, framing and conceptual motivations such as predictions and expectations (Smith, 2012). This is in line with the approach that stresses the interplay between how we use everyday visual perceptual abilities to make sense of movies and how movies have evolved to accommodate and conform to our existing cognitive faculties (Anderson, 1996; Berliner & Chen, 2011; Cutting, 2005, Smith, 2012). Therefore, studying movies teach us about our minds and the mechanisms we use to make sense of these complex visual stimuli.

We still don't know enough about how much spatial detail is retained across movie shots and which cues are essential for faster and more accurate spatial updating and integration of consecutive viewpoints in movies. Adherence to continuity editing rules deal mostly with spatial consistency, which suggests that maintaining relatively consistent spatial relations between agents and objects across shots is crucial to ensure that the viewer can allocate their limited attentional resources to following the narrative. As Kraft et al. (1991) put it, in film "the framework for coherence is spatial" (p. 603). In this respect, spatial continuity appears to be a given necessity that viewers grow to anticipate but do not pay very much attention to unless it is violated. So, the implied lack of attention to spatial details do not in fact reflect their trivial nature in narrative continuity but to the contrary it indicates our visual system's sensitivity to them and why most of the editing conventions are centered around those. One can say that it's because they are very crucial to narrative continuity that their consistency should be assured. This would in turn avoid confusion and disorientation in viewers and help them allocate their cognitive resources to following the content.

The main motivation of this dissertation is to provide a better understanding of how viewers form and maintain coherent spatial relations in movies where information needs to be integrated across consecutive viewpoints. The following series of experiments have inquired about the role of continuity editing rules on people's memory and judgments of agent-object relations in movies. The role of directional continuity and congruency of agent-related cues were examined with respect to accuracy and speed of spatial relation judgments.

The following chapters focused on four main lines of questions:

1) Does adherence to 180-degree rule affect the nature and the speed of viewers' spatial judgments with respect to the position of actors in a movie scene? How much do viewers base their spatial judgments on the establishing shot versus congruency of orientation? How robust is information for scene identity and orientation in long-term memory?

2) How much does congruency of agent-related cues (body position, gaze direction) affect viewers' judgments for future position? Also, how fast do viewers make shot-to shot transitions to decide about the projected position of an actor?

3) Do viewers correctly remember viewpoint-related information in shot-reverse shot sequences, which depict more than one actor's perspective? In relation, do viewers encode the relative positions of objects in movie shots according to actors' viewpoints?

4) Does viewing angle and emphasis on action affect perspective taking when viewers were asked to locate objects in a visual scene involving two agents? Does introduction to establishing shot affect people's perspective judgments?

CHAPTER 2

METHODOLOGICAL OVERVIEW

Experiments

Four groups of experiments were carried out for the dissertation, which will be explained in detail in the following chapters.

Experiment 1 investigated the role of continuity editing rules in maintaining congruent actor positions across movies shots. Experiment 1A used a 2 (establishing shot) X 3 (congruency) within-subjects design to examine how much viewers relied on the establishing shot versus spatial congruency to correctly identify the relative position of actors in a scene. It also investigated reaction times for those position judgments. Experiment 1B examined how well viewers retain scene-specific and orientation-related information about movie scenes in long-term memory. In a two-step procedure, this experiment investigated long-term memory for orientation and scene identity for movie scenes from Experiment 1A watched the previous day. Experiment 1C was conducted to get a baseline measure of preference for image orientation to assure that the mirror images are perceived similarly to the original orientation of a movie scene.

Experiment 2 inquired into viewers' prediction of future position for agents across movie shots in a 2 (cue order) X 2 (body side) X 3 (gaze side) within-subjects design. It examined whether viewers can correctly identify the upcoming position of an actor depending on the directional congruency of agent cues (gaze and body direction). It also examined reaction times for viewers' judgments with respect to the upcoming location of an actor.

Experiment 3 examined memory for individual shots in movie scenes where two actors' perspectives were depicted in a shot reverse shot sequence. Everyday activities involving two agents were filmed in multiple settings for this series of experiments. Experiment 3 considered viewers' memory for object locations with respect to each actor's perspective in regular viewing conditions versus when the axis of action was crossed. This experiment examined whether maintaining the 180-degree action line in a movie scene facilitates the recognition of object positions with respect to an actor's perspective.

Experiment 4 examined perspective taking for different viewing angles in movies, using photographs of visual scenes where two objects were positioned next to each other on a table. This experiment was carried out in a 3 (contact) X 2 (shot type) between-subjects design where open-ended responses were coded into different perspective categories. This series of experiments inquired whether the number of actors in a scene, contact with objects as well as instruction affected which perspective people adopted to describe object relations. Experiment 4A focused on how the number of agents in a visual scene (single actor vs. two actors) affects which perspective viewers adopt when describing object relations. In Experiment 4B, an action verb was used in the question to inquire about whether an action-related emphasis would alter the perspective judgments of the viewers. In Experiment 4C, a slight alteration in the methodology was implemented and participants were first introduced to an image showing the establishing shot of the scene before asking about the test image. The purpose of this manipulation was to further examine which frame of reference people adopt in visual scenes that approximate a movie-like scenario where the same scene is presented more than once from different angles.

All the experiments were conducted using MATLAB's graphical user interface (GUI) (https://www.mathworks.com) and stimuli were showed on a PC, which was run on Windows 7. For analyses, generalized estimating equations (GEE) under generalized linear models was used for binomial dependent variables to account for repeated subject effects, where Wald Chi-Square values were reported. Also, generalized linear mixed model (GLMM) was utilized for continuous dependent variables in within-subject designs, again to account for random subject effects. In addition, linear regression was employed for relationships between continuous variables and Pearson-Chi Square tests and multinomial regression were used for frequency analyses for multinomial dependent variables, unless otherwise indicated.

Participants

Participants were all Cornell University students who received course credit for their participation. Studies were approved by the Institutional Review Board (IRB) and all subjects signed a consent form before taking part in the studies.

Forty-two subjects (16 Men, 26 Women; 18-30 years old) participated in Experiment 1A and one subject was excluded from the results due to technical problems; 24 subjects participated in Experiments 1B (7 Males, 17 Females; 18-22 years old) and one subject was excluded from the results due to absence from the second part of the experiment and Experiment 1C was run as an addition to Experiment 3 and employed the same subjects. Forty-two subjects (11 Males, 31 Females; 18-23 years old) participated in Experiment 2 and three subjects were excluded from the results due to technical problems. Twenty-eight subjects (9 Men, 19 Women; 18-22 years old) participated in Experiment 3 and one subject was excluded from the results due to problems with the procedure. Experiments 4A through 4C were very brief and added to the beginning of other experiments, using the same subjects and each involved 56 subjects.

Terms

One major way to achieve spatial continuity in film is to use *180-degree rule*, which establishes the line of action (Bordwell & Thompson, 2003; Mercado, 2011; Smith, 2012). All shots in a scene are typically positioned on the same side of this line (Figure 1). Crossing this line can cause disorientation of the viewer, with respect to the relative position of objects and agents involved in the scene. The line of action changes within and across scenes by the usages of the establishing and re-establishing shots, or through the movement of the camera and actors.

180-degree violation typically occurs when the camera crosses to the opposite side of the action line, which in turn changes the background of the subsequent shot and reverses the relative position of actors. 180-degree violation can also be simulated after the fact by *flipping* shots 180 degrees horizontally (*mirroring*) to reverse the relative position of the actors while keeping the background constant. For the purposes of this dissertation, the former will be referred to as the *actual 180-degree violation*, while the latter will be referred to as the *simulated 180-degree violation.*

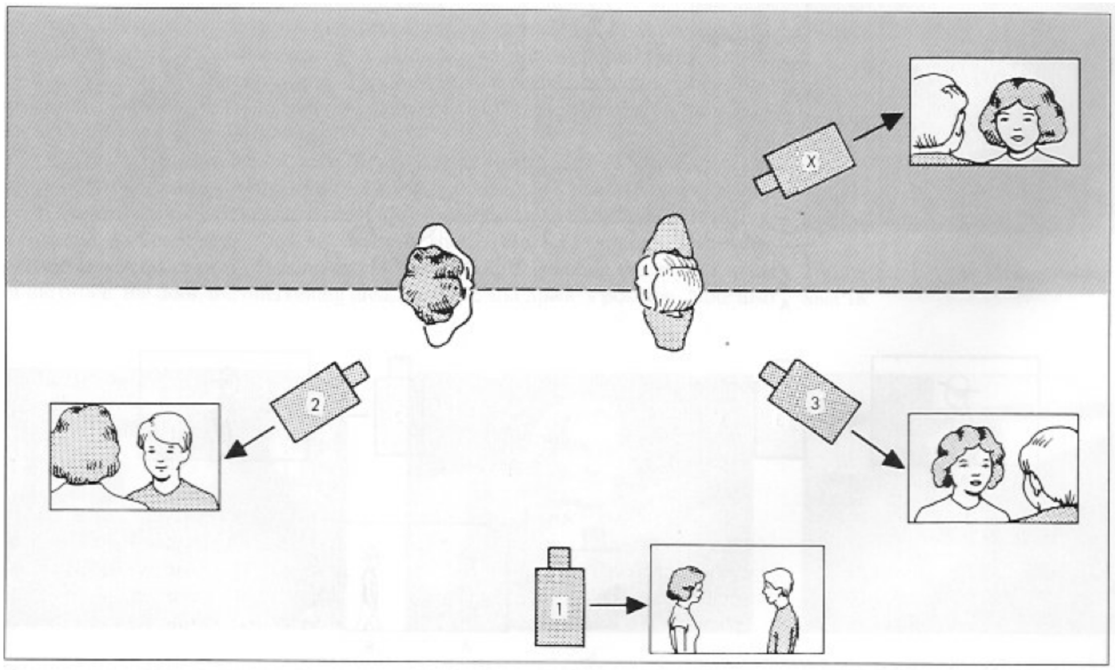


Figure 1. 180-degree rule and the establishment of the axis of action. The above image depicts the employment of the 180-degree rule and the placement of cameras in a typical scene. The horizontal line indicates the line of action. To adhere to the rule, all the cameras (1,2,3) should be placed on the same side of the 180-degree line. In this example, cameras 2 and 3 show examples of shot reverse shots, while camera 1 shows the establishing shot for that scene, all of which are defined below. Here, violating the 180-degree rule would mean to shoot with the camera labeled "x", which is positioned on the opposite side of the action line (Bordwell & Thompson, 2003, p. 263).

A scene generally starts with long shots, which are called *establishing shots*. Those are usually *long or extreme long shots* with wider angles that present the layout of the scene and relative positions of actors and objects involved (Bordwell & Thompson; Chandler, 2009; Mercado, 2011). A long shot refers to a shot, which is taken from a distance and depicts a wider visual angle. The establishing shots are usually objective (third person) shots. An *objective angle* shot is defined as a "sideline viewpoint" in which the audience experiences the film from none of the actors' point of view (Mascelli, 1965). *Close-ups* were then used in conversations, which employ mostly *shot-reverse shots* and *gaze matches* to promote spatial continuity (Bordwell &

Thompson, 2003, Magliano & Zacks, 2011).

A *shot reverse shot* (SRS) sequence is the editing technique usually used in conversations. In a SRS sequence, the relative position of each actor is consistent and the conversation partners appear at the opposite sides of the screen, maintaining the 180-degree action line (Bordwell, 1985; Chandler, 2009; Smith, 2012). SRS sequences employ either *over the shoulder* (OTS) shots, or off-screen gaze matches. The *gaze matches*, also called *eye-line matches* are produced by cutting to where the actor was looking in the previous shot to create a continuity in their gaze (Bordwell & Thompson, 2003; Chandler, 2009; Smith, 2012). In OTS sequences, one of the actors faces the camera and is in focus, while the other actor's back (usually up until the shoulders) is in the foreground. In these cases, the viewers get to see both actors but only one's perspective is emphasized. In SRS sequences that rely on gaze matches, only one of the actors is visible in each shot, looking off the screen to the conversational partner, slightly at an angle.

A point of view (POV) shot is a special case that has been defined differently among cinematographers as to what it exactly entails. While it is sometimes defined as depicting the optical point of view of a character (Chandler, 2009), a POV shot assumed different descriptions among cinematographers and researchers who study film. It is commonly defined as the shot resulting from cutting to what an actor is looking at off screen in the previous shot to create an optical continuity in gaze, without necessarily showing the exact angle of the character's gaze (Bordwell & Thompson, 2003; Smith, 2012). Cinematographers also defined POV shots along a continuum of subjectivity and objectivity where it fell closer to the subjective dimension without assuming the characteristics of a purely *subjective* shot, which is also called the *subjective camera.* A subjective shot refers to a shot where the actor looks directly into camera (Mercado, 2011). This can be uncomfortable if used for a long duration and can lead to breaking the illusion of film (4th wall) if the actor speaks to the audience. In a subjective shot therefore, the viewer exchanges places with the camera. All other shots in that regard are versions of objective shots where the viewer is in the position of an observer. Mascelli (1965) in his book *Five C's of Cinematography* defined this relationship as the following: "A point-of-view shot is as close as an objective shot can approach a subjective shot—and still remain objective. (…) The viewer does not see the event through the player's eyes, as in a subjective shot in which the camera trades places with the screen player. He sees the event from the player's viewpoint, as if standing alongside him. Thus, the camera angle remains objective, since it is an unseen observer not involved in the action" (page 22).

A *viewpoint* has a broad definition and refers to the vantage point where a scene is observed while *perspective* and *point of view* are usually used to describe a more subjective viewpoint, which belongs to a person, actor or character (Franklin, Tversky & Coon 1992; Tversky, 2004). In this dissertation, perspective will be used to refer to a viewer's or actor/camera's viewpoint and point of view will be used to refer to a type of shooting technique.

CHAPTER 3

EXPERIMENT 1: SPATIAL CONGRUENCY AND RELATIVE POSITIONS OF AGENTS IN MOVIE SEQUENCES

Experiment 1A

We know from previous literature that people can successfully follow successive shots in a movie rather easily. The main motivation of the following experiment was to further understand how viewers form coherent visual maps in movie scenes out of a series of shots taken from separate viewing angles and make shot transitions fast enough to follow the narrative.

Film editors employ several procedures to maintain spatial continuity in movie sequences. While one of these techniques is the adherence to the 180-degree action line, the other important one is the employment of the establishing shots, which depict the relative positions of actors and objects from a wider angle (Bordwell & Thompson, 2003; Kraft et al., 1991; Mercado, 2011; Smith, 2012). In an early study, Kraft et al. (1991) used slideshow of drawings showing an interaction between two actors. They found that recognition memory was higher for shots where 180-degree rule was maintained. Also, when asked to recreate the scene, a reliance on establishing shots was observed. Those shots are suggested to establish the foundation for spatial relations in a scene. One of the motivations of this experiment is to look at actor relations in more complex scenes where the viewers need to track the relative positions of multiple actors across shots. Also, in contrast to Kraft et al. (1991), actual movie clips were used to increase the validity of the stimuli. This experiment also used judgments of the relative positions of actors as test stimuli, which called for a more active role on the part of the viewer to integrate information than simply remembering shot instances.

Most of previous research with dynamic scenes used either simple animations, or simple interactions between actors in lab settings instead of mainstream movies. In such a study, Huff and Schwan (2012) used simple animations of moving objects (i.e. a car on a road) to investigate how screen direction might affect spatial updating. The researchers found that people were more accurate and faster in their judgments of motion direction when the 180-degree rule was maintained even when the angular distance of the object between shots was large, compared to when the angular distance was smaller but the screen direction was incongruent. The authors proposed that people use screen direction rule as a heuristic strategy for spatial updating across angle changes. They suggested that people keep track of screen position of objects instead of a more effortful updating based on spatial alignment. Under this spatial heuristic model, the violation of the axis of action increases reaction times because when people can't use the screen direction strategy, which saves time, they resort back to spatial alignment to compute the distance. This results in longer reaction times when the angular distance increases.

Previous studies also showed that people were good at detecting screen direction violations (64-70%), however those mostly inquired about the detection of the violation on change detection not its effects on forming accurate judgements for actor positions (Baker & Levin, 2015). As discussed above, Baker and Levin (2015) found that changes in object properties (color of an actor's clothing) were detected the most when they coincided with the violation of the 180-degree rule. The authors suggested that the direction violation was salient enough to be detected and attract attention to the shot when the violation occurred, which in turn led to the detection of the change. This again shows that adherence to the 180-degree rule is a strategy to both hide the cut and the increase the smooth integration of information between shots. When a spatial violation happens, it is salient enough to attract attention from the narrative and towards the perceptual details.

One of the lesser-studied aspects of movies is the relative position of actors in a scene and how viewers keep track of them. Studying how spatial continuity rules affect the formation of spatial maps for actor positions is crucial to discover more about how viewers achieve shot integration with ease. The broader motivation here is to understand what viewers rely on the most to understand the relative positions of actors. Do they depend on shot to shot consistencies or does the establishing shot provide them with the necessary information? In relation, do viewers still recognize which establishing shot belongs to a certain scene when exposed only to individual shots? Another rather understudied aspect of movie scenes is the complexity with respect to number of actors in a scene. The question is whether continuity editing rules are robust enough to make that complexity irrelevant.

With these in mind, the Experiment 1A intended to answer the following questions: Does adherence to 180-degree rule affect the nature and the speed of viewers' spatial judgments with respect to the position of actors in a movie scene? How much do viewers base their spatial judgments on the establishing shot versus congruency of orientation?

The initial expectation was that when the scene includes an establishing shot, people would rely on it more when the 180-degree rule was violated. In the absence of establishing shots, people would base their spatial judgments on the congruency of orientation based on the 180-degree rule. When this is violated, they could resort to the information presented in last shot, which would be the most efficient comparison, especially in the case of multiple actors.

Methods

Stimuli. The study material consisted of short video clips (~1 min long) taken out of a subset (see Filmography) of a larger Hollywood movie sample used in a former study (Cutting, Brunick & Candan, 2012). The same stimuli were used for Experiments 1A through 1C. While the examples of these stimuli shown in this dissertation reflect the original aspect ratios for these selected movies, the experiments employed resized (256 X 256) versions of the clips similarly to the original stimuli used in previous studies (Cutting, DeLong & Nothelfer, 2010; Cutting, Brunick & Candan, 2012).

In Experiment 1A, a 2 (establishing shot) X 3 (congruency) within-subjects design is employed. The clips were presented either with (*present*) or without (*absent*) the establishing shot. In each clip, the shots were either presented in their *original* version (*congruent*), where all the shots adhered to the 180-degree rule, or some of the shots were altered to violate the 180-degree rule. The violation was applied either to the middle shot (*middle shot flipped*) or to all the shots in the second half of the clip (*second half flipped*), depending on the condition. The violation was achieved by flipping the video horizontally, which resulted in the orientation of the actors to be switched by 180 degrees (Figure 2). This experiment employed a *simulated 180-degree violation*. As explained in the previous section, in a simulated violation, the relative position of the actors and objects are reversed when the shot flips 180 degrees. This also flips the background on the screen but the content of the background remains the same as opposed to an actual 180-degree violation. That one is shot from the other side of the 180-degree line, which features an altogether different background. Also, in the experiment scenes, the actors remained stationary and did not change their relative positions.

Figure 2. Example for 180-degree violation used in Experiment 1A. These frames are taken from one of the clips used from the movie *Erin Brockovich* (2000). The images are shown in the movie's original aspect ratio (1.85:1). The images exemplify how flipping the above-left shot 180 degrees horizontally simulated the 180-degree violation, by reversing the relative positions of actors.

For each movie, two scenes were chosen to reflect different levels of *complexity* mainly based on number of actors. Those scenes were coded with respect to the following five low-level features: *screen duration, number of shots, average shot duration, number of actors and number of camera positions* (Table 1).

Screen duration was defined as the total duration of a movie scene in minutes and the number of shots was defined as the total number of shots present in a scene. Average shot duration (ASD) was calculated as the mean shot duration in seconds for each clip as a factor of screen duration and number of shots. The number of actors (NOA) was defined as total number of actors who were visible at any time in the scene. Number of camera positions (NOC) was defined as the total number of unique positions the camera was placed within a scene. This was coded considering both the *angle* and *distance* of the camera (Figure 3 and Figure 4). If the same angle and distance combination was used more than once, it was considered one unique position. Therefore, the NOC variable considered the total number of unique camera positions employed in a scene.

Movie name and year	**Scene no**	**Screen duration (min)**	**Number of shots (NOS)**	**Average shot duration (ASD)**	**Number of actors (NOA)**	**Number of camera positions (NOC)**	**Mean Reaction Time (s)**
Erin Brockovich (2000)	1	1.42	17	6	2	4	2.69
	2	1.09	17	4.06	6	7	2.78
Five Easy Pieces (1970)	1	0.58	17	3.41	2	5	2.55
	2	1.21	21	3.86	7	7	5.22
Ordinary People (1980)	1	1.07	13	5.15	2	4	2.91
	2	0.53	15	3.53	3	5	2.84
Social Network (2010)	1	1.10	25	2.8	4	8	3.69
	2	1.10	33	2	13	13	7.76
Valentine's Day (2010)	1	0.48	13	3.69	2	2	3.48
	2	0.52	13	4	3	4	2.35
What Women Want (2000)	1	1.03	23	2.74	2	4	3.51
	2	1.01	17	3.58	8	8	3.80

Table 1. Descriptive figures for scene properties in Experiment 1A. The above table shows descriptive figures for scene duration, number of shots, average shot duration, number of people and camera positions with RT measures for each scene used in the sample.

Figure 3. Examples for camera position with respect to distance change. The above shot pairs use the same angle, but a closer distance (close-up) in the second shot focuses on one actor. This in turn increases the size of the actor on the screen, while restricting the visible background and limiting information for the relative position of actors close-by. The above shot pair is taken from the movie *Social Network* (2010) and the below pair is taken from the movie *Five Easy Pieces* (1970).

Figure 4. Examples for camera position with respect to angle change. The above shot pairs use different camera angles: frames on the left are taken approximately 45 degrees on right side of the axis of action and frames on the right are taken approximately 45 degrees on the left side of the axis of action. The above shot pair is taken from the movie *Social Network* (2010) and the below pair is taken from the movie *Five Easy Pieces* (1970).

Procedure. During the experiment phase, the participants watched 12 video clips on a computer screen. Each subject was exposed to all experimental conditions in a random order and condition-scene pairs were counterbalanced across subjects. After viewing each clip, participants were asked to perform a 2AFC (two-alternative forced choice) in response to the following question: “Which image better represents the position of actors in the previous clip?" In each trial, they were presented with both the *original* establishing shot, which shows the relative positions of the actors in the original movie and the *mirror* image of the establishing shot, where the relative position of the actors was reversed by flipping the image 180 degrees (Figure 5). Those were randomized with respect to location on the screen (right vs. left). Reaction time measures were also recorded in addition to the decision measure.

Figure 5. Examples for test stimuli used in Experiment 1A. The above image pairs exemplify the test images used the decision task. The above image pair shows the original frame (on the left) vs. mirror frame (on the right) from the establishing shot of the movie *Ordinary People* (1980). The below image pair shows the original frame (on the left) vs. mirror frame (on the right) from the establishing shot of the movie *Five Easy Pieces* (1970).

After the completion of the experiment, the participants were given a post-experiment questionnaire where they were first asked to indicate whether they have previously watched any of the movies used in the experiment. This questionnaire also employed open-ended questions similar to the ones used in previous studies on *change blindness* (Angelone et al. 2003; Levin & Simons, 1997; Simons & Levin, 1998). Participants responded to a set of three open-ended funneling-type questions to target the awareness of the video manipulation with respect to the 180-degree violation. The questions were the following: Q1: "Did you notice anything unusual about the short video clips you've just watched? If yes, please explain your response"; Q2: "Do you think that the position of the actors was congruent in all the shots you've watched? If no, please explain your response"; Q3: "Did you notice that in some of the shots, the orientation of the video was flipped so that the actor positions were changed 180 degrees? If yes, please explain your response". The questions were concealed with a white paper sheet and were revealed to the subjects one at a time, as they finished answering the previous question.

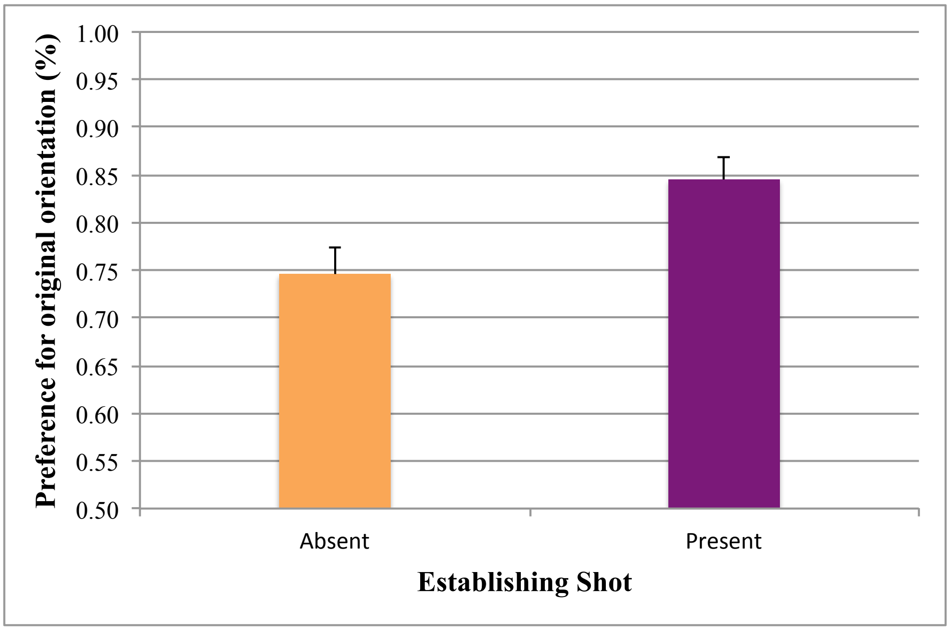
Previous research on change detection in motion pictures found that only 1 out of 10 participants noticed changes to peripheral details across movie shots and this only increases to 33% for more central features like the main actor who changes to another one in a subsequent shot (Angelone et al. 2003; Levin & Simons, 1997).

In a more recent study, Baker and Levin (2015) found higher detection rates (68%) for object properties across shots when those coincided with a 180-degree violation compared to when they did not (48%). They also showed that people were good at detecting the 180-degree violation when it was accompanied with object

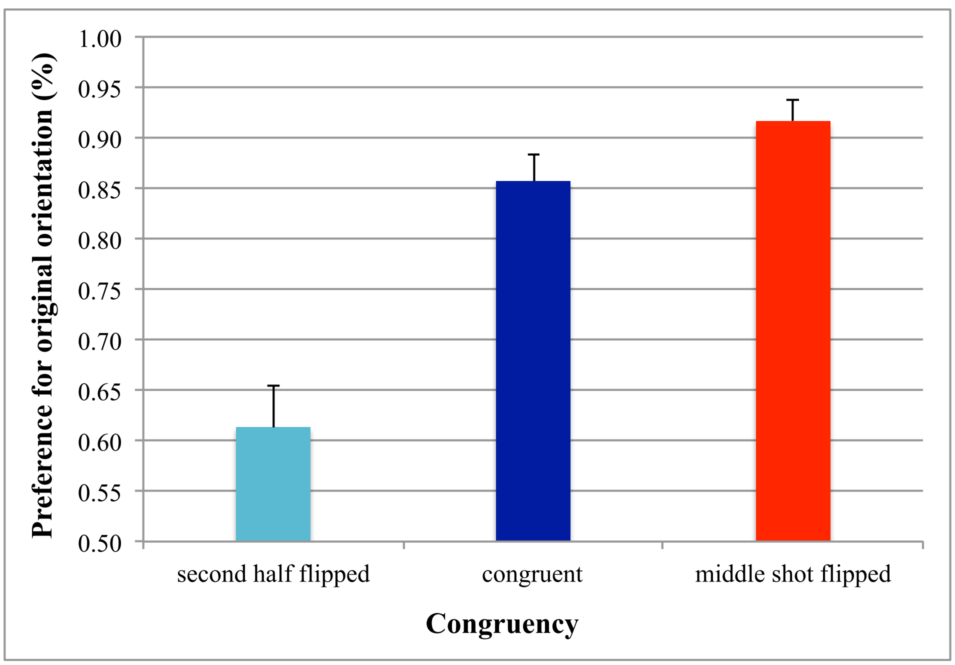
property changes (64%) compared to when it was by itself (46%). This number increased to 70% when the background remained unchanged. Detection of the 180-degree violation is generally explained by this jarring change in relational properties, which might potentially make the cut more detectable to the viewer, therefore attracting the attention to the details of that shot.
(Please refer to the *Participants* section in Chapter 2 for detailed information on subjects used in each experiment).

Results

For the analyses, preference for the original orientation of the establishing shot was considered as the dependent variable. Overall, there was an effect for congruency on preference (χ^2 (2, N=42) = 40.88, p<.001). Post-hoc analyses with Bonferroni correction indicated that there was a difference in performance for the second-half flipped condition (p<.0001) compared to the congruent and middle-shot flipped conditions, which were similar (p=.35) (Figure 6). This indicated that viewers chose the original orientation more often when the 180-degree rule was preserved throughout the scene and when only the middle shot was flipped. However, when both halves of the scene displayed reversed spatial relations, viewers' preference for the original orientation decreased significantly. An effect for the presence of the establishing shot was also observed (χ^2 (1, N=42) = 5.69, p=.017), which indicated that preference for the original orientation was higher when an establishing shot was present in the beginning of the scene compared to when it was absent (Figure 6).



a)



b)

Figure 6. Mean preference for the original orientation with respect to establishing shot and congruence. **a)** This figure shows that the viewers' preference (%) for the original orientation was higher when the establishing shot was present in the beginning of the scene compared to when it was absent. **b)** This figure indicates that, viewers' preference (%) for the original orientation was lowest (61%) for the second half flipped condition, where the shots in second half of the scene were flipped 180 degrees so that the relations were reversed compared to the first half of the scene. Error bars represent one standard error of the mean.

Interestingly, the presence of the establishing shot had different effects depending on congruency manipulation (Figure 7). For the second half flipped condition, viewers did not have a reliable preference for either orientation (original vs. flipped) (%50) when the establishing shot was absent but when the establishing shot was present, the preference for the original orientation increased to 70%. For the congruent condition, viewers preferred the original orientation even in the absence of the establishing shot (80%) but the preference was slightly higher when the establishing shot was present. Surprisingly, the presence of the establishing shot was not effective for the middle shot flipped condition (p=.988). It can be suggested that the orientation manipulation halfway into the scene and the return to the original configuration in the second half might have heightened the attention of the viewers, leading them to attend more to the second half of the scene, expecting another violation. This in turn, could have decreased the reliance on the establishing shot.

Looking at the characteristics of the movies, moderate effects were found for ASD (χ^2 (1, N=42) = 4.79, p=.029) and screen duration (χ^2 (1, N=42) = 4.90, p=.027) on preference. As we know that there is an inverse correlation between ASD and number of shots, for scenes with lower ASDs, people would need to consider more shots before deciding on the position of actors in the scene. This may have affected people's preference for the orientation of the establishing shot.

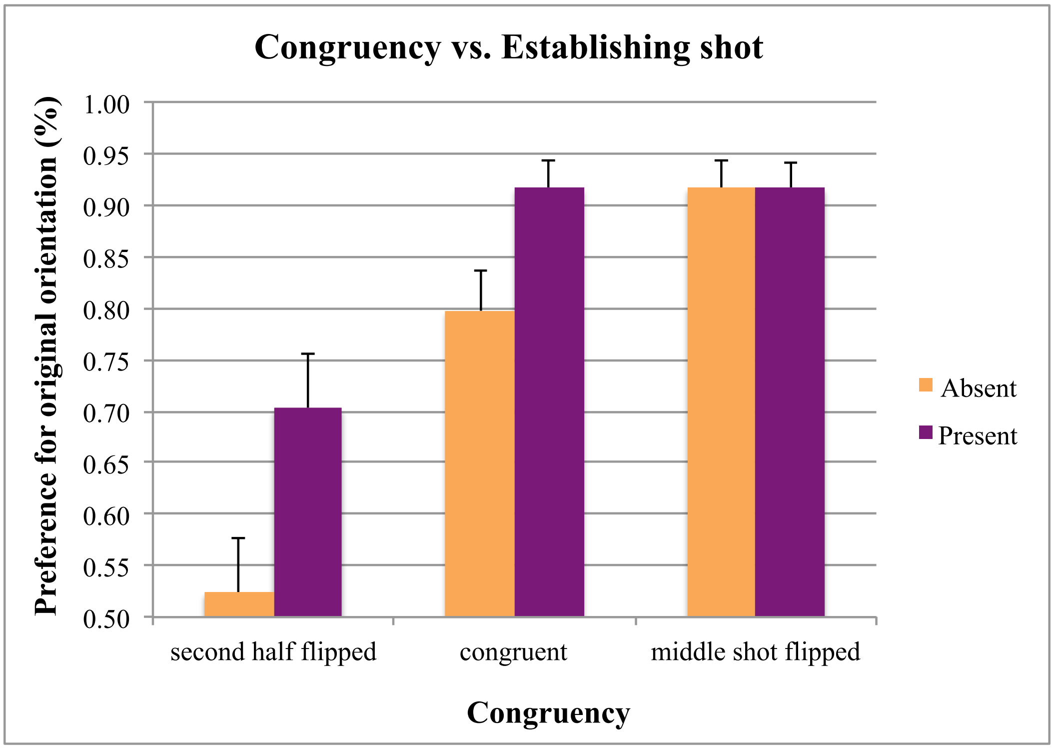


Figure 7. Intreaction between congruency and establishing shot. This figure indicates that the presence of the establishing shot had an effect on viewers' preference for the second half flipped and congruent conditions but did not have an effect for the middle shot flipped condition. For the second half flipped condition, viewers did not have a reliable preference for the orientation in either half of the clip (%52) when an establishing shot was absent but preference for the original orientation (first half of the clip) was higher when an establishing shot was present in the beginning of the scene. For the congruent condition, preference for the original orientation was slightly lower when the establishing shot was absent (80%) compared to when it was present (92%). Error bars represent one standard error of the mean.

With respect to reaction time measures, there was an effect for the establishing shot ($F(1, 452) = 10.17$; p=.002), indicating that people were slower in the orientation task when the establishing shot was absent in the beginning of a scene (Figure 8). Also, a modest effect was found for congruency ($F(1, 454) = 10.16$; p=.048), indicating that in the congruent condition, people responded faster to the orientation task compared to the second-half flipped and middle-shot flipped conditions, which were similar (Figure 9).

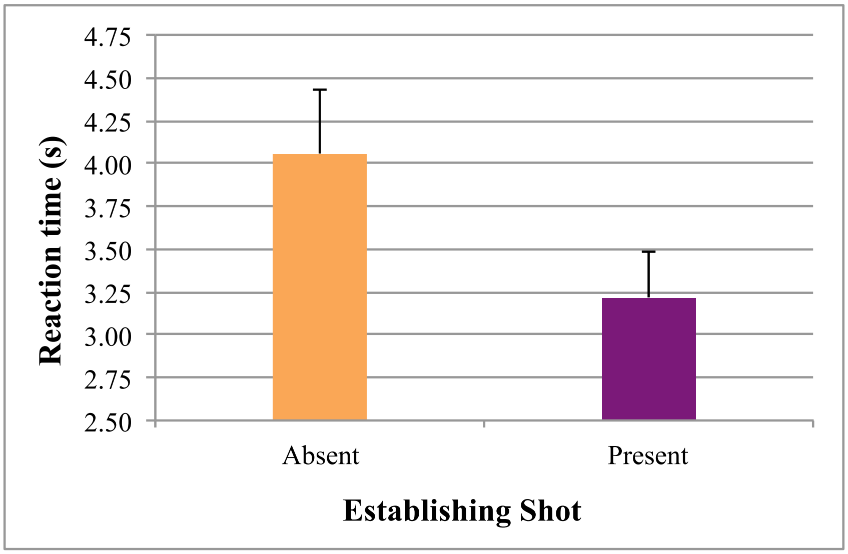


Figure 8. Mean reaction time results for establishing shot in Experiment 1A. The figure indicates that viewers' reaction times in the decision task were longer when the establishing shot was absent in the beginning of a scene versus when it was present. Error bars represent one standard error of the mean.

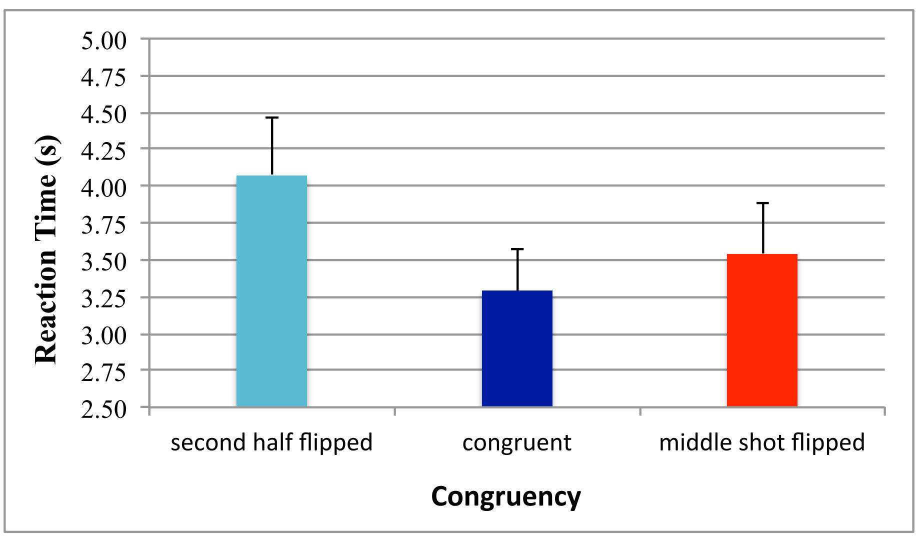


Figure 9. Mean reaction time results for congruency in Experiment 1A. Error bars represent one standard error of the mean. The figure indicates that viewer’s reaction times in the decision task was longest for the second half flipped condition and shortest for the congruent condition.

A major effect observed on reaction times was the difference across movies ($F(5,454) = 10.96$, $p < .0001$) (Figure 10). Post-hoc tests with Bonferroni correction indicated that the reaction times for all the movies were similar except for *Social Network* ($p = .002$ for *Five Easy Pieces* and for the remaining movies $p < .001$), where subjects took a significantly longer time to respond to the decision task.

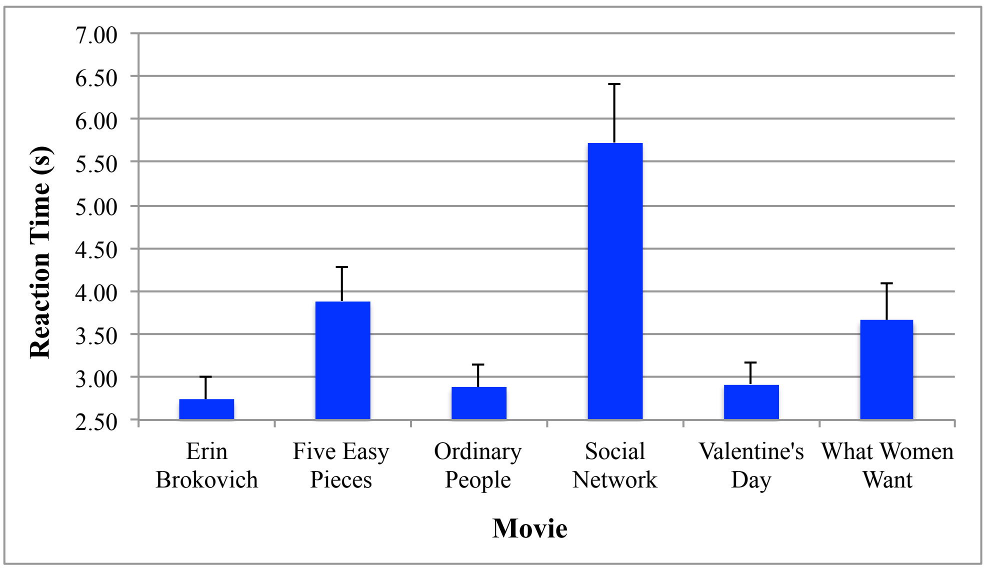


Figure 10. Mean reaction time results for movies. The above graph shows mean reaction time measures in the decision task for different movies. This figure reflects overall results aggregated for both scenes from each movie. Error bars represent one standard error of the mean.

With respect to our stimuli, *Social Network* is one of the more recent movies, with a greater number of shots compared to the other movies and had on average more actors (Figure 11) (refer to Table 1 for the mean reaction times for all the scenes in the sample). We know from previous research (Cutting, Brunick, DeLong, Iricinschi & Candan, 2011; Cutting, 2016) that the ASD has become shorter in more recent Hollywood movies. While there was negative correlation between ASD and reaction times (r (504) = -.225, p <.001, R^2 =0.051) (Figure 12), it did not reliably predict the reaction time results when considered together with the other characteristics of those scenes, mainly the number of people and camera positions.

Figure 11. Example shots from the movie *Social Network* (2000). The figure shows the establishing shot of one of the scenes (on the left) from the movie *Social Network* (2000) and an example shot, which shows the complexity of the scene with respect to number of actors.

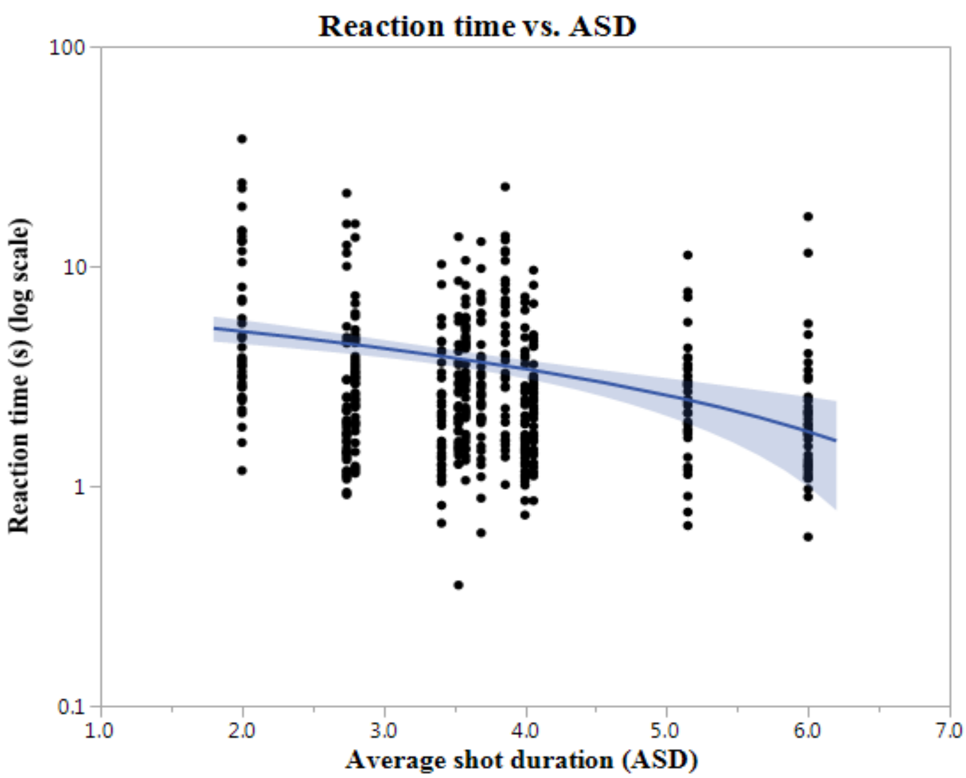


Figure 12. The scatterplot for reaction time vs. average shot duration (ASD) for movies used in Experiment 1A. The trend shows that as the ASD of a scene gets shorter, reaction time measures for viewers get longer.

Mixed linear regression analyses considering all movie characteristics indicated that there was significant relationship between the reaction times of viewers and the number of actors ($F(1, 454) = 21.874$; $p<.0001$) as well as camera positions ($F(1, 454) = 7.85$; $p=.005$) in a scene (Figure 13). As the number of actors in a scene increased, people's reaction times got longer. Also, a strong positive correlation between number of actors and number of camera positions (r (504) = .927, p <.001, R^2 =.86) (Figure 14) was observed, which indicated that as number of actors increased, the number of unique camera positions also increased (refer to Table 2 for the correlation figures between the all the movie characteristics).

	Screen duration (min)	**NOS**	**ASD (s)**	**NOA**	**NOC**
Screen duration (min)	1	.461**	.255**	.335**	.401**
NOS	.461**	1	-.674**	.695**	.817**
ASD (s)	.255**	-.674**	1	-.498**	-.562**
NOA	.335**	.695**	-.498**	1	.927**
NOC	.401**	.817**	-.562**	.927**	1

Table 2. Correlation between scene properties in Experiment 1A. The above table shows the correlation between screen duration, number of shots, average shot duration (ASD), number of actors (NOA) and number of camera angles for movie scenes used in the Experiment 1A. ** indicates significance at the .01 level. N=504.

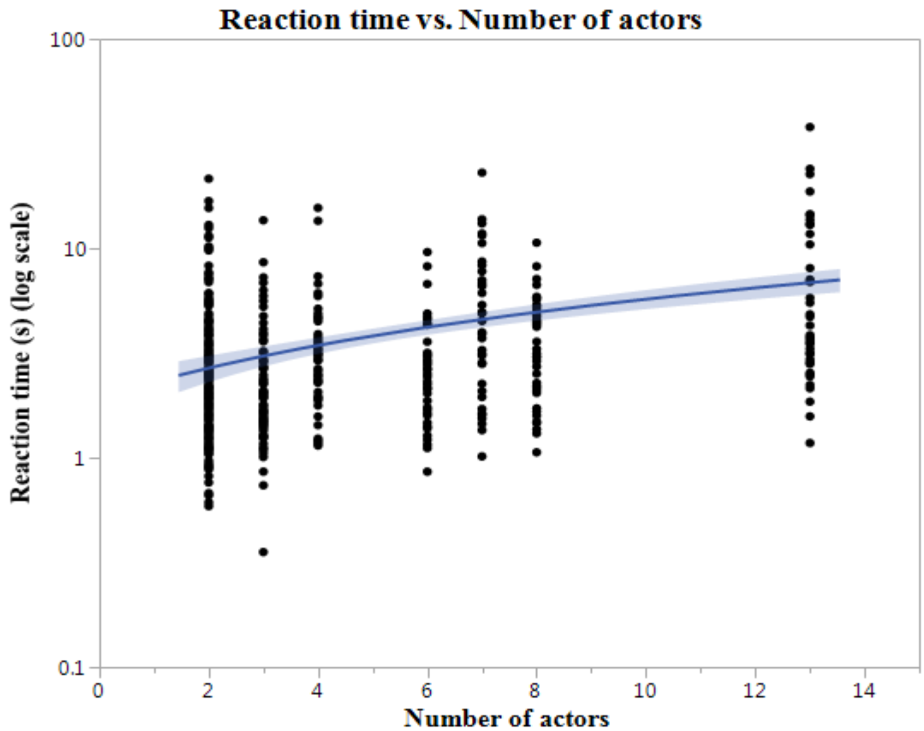


Figure 13. Reaction time measures with respect to number of actors in Experiment 1A. The above figure shows that reaction times increase as the number of actors in a scene increase.

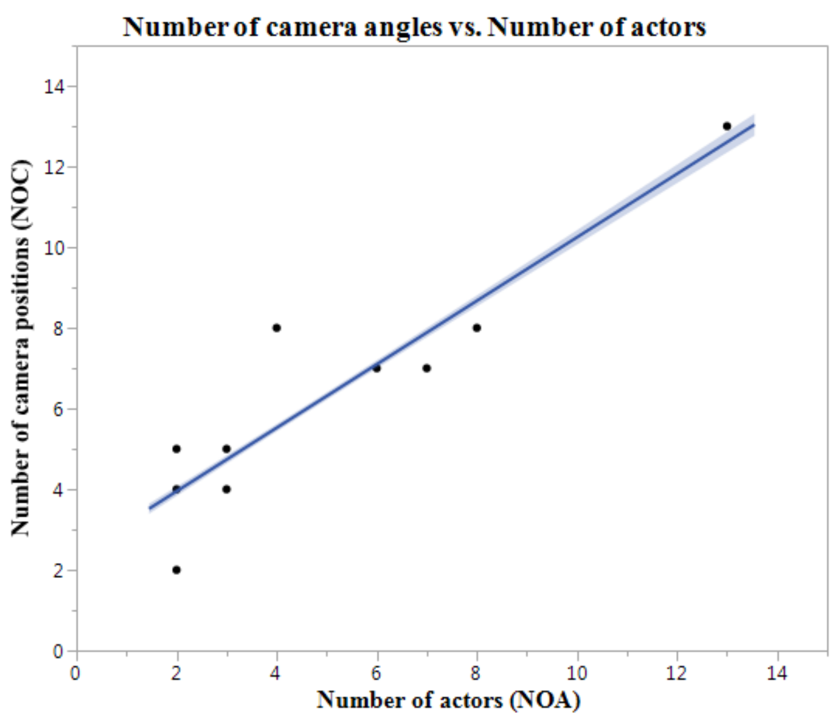


Figure 14. Scatterplot for correlation between the number of actors and camera positions for movie scenes used in Experiment 1A.

As part of the post-experiment questionnaire, subjects were also asked whether they had previously watched the movies used in the experiment. While only a small percentage had seen four of the movies (*Five Easy Pieces* (0%); *Ordinary People* (2%); *Erin Brockovich* (5%); *What Women Want* (9%); *Valentine's Day* (28%)); one, *Social Network* was seen by about half of the subjects (52%) before taking part in the experiment. Also, the post-experiment questionnaire results indicated that 78% of the subjects responded that they were aware of the manipulation (i.e. "Yes, I did notice the orientation flip in some of the shots, as before the change, one person would be positioned to the right of the other actor and later the same person would appear to their left"). Surprisingly, a similar percentage (76%) mentioned this awareness in the first question without being asked directly about the manipulation, although most of the time they did not provide much detail. This suggests that subjects were sensitive to the violation of the 180-degree rule and most them noticed the change of orientation within clips.

The results from the post-experiment questionnaire are in line with what we would expect based on earlier research that showed that people were sensitive to 180-degree violations in movies when asked directly and were better at recognizing a feature change in an object if it coincided with that violation (Baker & Levin, 2015). In this experiment, it was not possible however to differentiate whether participants were only referring to the conditions in which all the shots in the second half were flipped or they have also noticed the manipulation when it was applied to the middle shot only. Since there were two conditions that were manipulated for adherence to 180-degree rule, one being only for the middle shot and the other for the whole second part of the video, further research can determine whether people can still realize an orientation violation when only one shot within a clip was reversed. Also, these results reflect people's judgments after being exposed to all the conditions in the experiment. In future research, it may be interesting to examine the initial detection of this violation with an online detection task.

Discussion

Overall, the results from Experiment 1A indicated that congruency as well as the presence of the establishing shot affected people's judgments for actor positions in a movie scene. People relied on the orientation of the shots to decisde on actor positions even in the absence of the establsihing shot, as long as the spatial relations were consistent throughout the scene. When both haves of the scene displayed opposite rletaions, viewers did not show a particular preference for either half, especially in the absence of the establishing shot, indicating that they did not simply rely on the last encounterred information. This suggests that both of the cues were informative in people's decision and people treated the scene as a whole and did not only depend on the last observed orientation to indicate their decision for the position of actors in a movie scene. When no establishing shot was present and the two parts of the scene showed different relations, people did not necessarily rely on the second part of the scene as suspected but they were equally likely to refer to either part for their judgment. This suggests that consistency of orientation throughout a movie scene, in this case adherence to the 180-degree rule, is important for unified consistent spatial judgements for actor positions.

Also interestingly, preference was not affected when there was only one instance of the direction violation applied to the middle shot. This was apparent from the fact that the presence of the establishing shot affected the rest of the conditions except the middle-shot flipped condition. This suggests that in those scenes, attention might have been hightened due to the motion transient associated with the directional violation, which might have made the cut more detectable. We can speculate that this might have influenced viewers to pay more attention to the second half of the clip, possibly expecting another violation. In that regard, they may have relied less on the establishing shot but more on the relations in the second half of the scene.

In addition, ASD and screen duration also affected spatial judgments. As we know that there is an inverse relationship between the ASD and number shots, in scenes with lower ASDs, people would need to consider more shots before deciding on the relative positions of actors. Also, the duration of a scene affects how much time passes after one is exposed to an establishing shot. So, duration effects might be related to a decay in memory for information related to the establishing shot. This also relates to the underlying motivation for the employment of what is called *re-establishing* shots in movies. Re-establishing shots are versions of establishing shots taken from a wider third person perspective to either remind the viewers of the existing relations after elapsed time or to redefine the relative positions of actors when those change due to movement, introduction a new character or a new action line.

Similar effects were also present in reaction time measures, people were slower to respond to scenes where the establishing shot was absent and when the shots were flipped for the second half of the scene. This suggests that viewers did not readily select the orientation of the most recent shot but needed more time to mentally transform the relative positions of actors when both halves of the scene had conflicting information. Reaction times were also correlated with movie characteristics, especially the number of people present in the scene, which was also closely associated with the number of camera positions used. This suggests that people needed more time to integrate information into a spatial map when there were more relations to consider, making a scene more complex. Also, scenes that involve multiple people tended to use more close-ups, which also limit the spatial cues on the screen with respect to the relative position of actors. This could be one reason why scenes with multiple people might employ additional spatial cues like gaze and head turns to facilitate the spatial continuity and narrative comprehension across shots.

For Experiment 1A, the scenes were chosen from situations where the actors were stationary so their relative positions didn't change within the sequence. For a follow-up, it would be important to also consider sequences in which actors move to change their positions within a sequence and examine whether the viewers can determine the actors' end positions relative to their initial positions. Finallly, the post-experiment questionnaire results indicated that most the subjects were aware of the manipulation. This again indicated their sensitivity to violations of spatial continuity in movies.

Experiment 1B

Experiment 1A showed that the presence of an establishing shot and the congruency between consecutive shots affected the pace and the nature of people's spatial judgments with respect to actor positions. This follow-up experiment inquired how persistent these effects were long-term. Previous research showed that long-term memory (LTM) for still images is very robust (Brady, Konkle, Alvarez, & Oliva, 2008; Konkle, Brady, Alvarez, & Oliva, 2010; Standing, 1973). While this might be the case, fewer studies investigated persistence of memory in the case of dynamic scenes (Matthews, Benjamin & Osborne, 2007; Meyerhoff & Huff, 2016) as well as how much orientation-related information is preserved over time (Joubert & Oliva, 2010; Kwok & Macaluso, 2015). While some research showed above chance performance (72%) for LTM fidelity with respect to image orientation (original vs. mirror foils) (Joubert & Oliva, 2010), little is known about how well orientation-related information persists over time for movie scenes.

Also, another aspect that makes the investigation of movies different than still images is the fact that the viewing activity of a movie has an inherent long-term nature, meaning that viewers periodically encounter the same characters engaging in different activities in similar settings throughout a movie. How well do people differentiate between different encounters in similar settings over the course of a movie? This question relates to the research on *gist* versus *specificity* of visual scenes that was researched for a long time with still images (Goldzieher, Andrews, & Harris, 2017; Greene & Oliva, 2009; Oliva, & Torralba, 2006). The question remains about how well viewers retain scene-specific as well as orientation-related information about movie scenes over time. The specific questions inquired in the following experiment were the following: Do people remember the position of people in a movie scene after a delay of one day? How well can viewers identify the scene they have viewed compared to similar alternative scene?

Methods

Stimuli and Procedure

Experiment 1B was an extension of Experiment 1A where participants took part in a LTM test. This was a two-step experiment where participants first viewed the clips on the first day and performed the orientation task and were asked to come back to the lab the next day (at the same time) to take a follow-up memory test. In day two of the experiment, they were asked to perform a 4AFC (four-alternative forced choice) task using the stills from the establishing shots. Subjects were asked: "Which image correctly shows the scene AND the position of actors for the clips you have watched yesterday?" This task was used to test for scene *identity* (original vs. alternative) and scene *orientation* (original vs. mirror) (Figure 15). The test trials were randomized for the order of scene and movie as well as the position of the images on the screen. Reaction time measures were also recorded in addition to the decision.

a) original scene/ original orientation **b)** original scene/ mirror orientation

c) alternative scene/ original orientation **d)** alternative scene/ mirror orientation

Figure 15. Example test stimuli used in Experiment 1B. The above figure exemplifies one of the test trials in the 4AFC decision task, which asked viewers to identify the correct scene and orientation (image on the top left corner) for the establishing shot that belonged to one of the clips taken from the movie *Five Easy Pieces (1970).*

Results

The responses from the 4AFC task were analyzed for accuracy with respect to two dependent variables: scene orientation and scene identity. First, a cross-tabulation was done to compare the individual observations for the orientation judgments from day 1 with the orientation and identity judgments from day 2 (Table 3). There was a strong correlation between orientation judgments on both days (χ^2 (1) = 28.61, p<.001), and a moderate correlation between orientation and scene judgments on the second day of the experiment (χ^2 (1) = 5.65, p=.017). This suggested that orientation accuracy from day 1 accurately predicted the orientation accuracy on day 2 and people who correctly identified the orientation were more likely to correctly identify the scene and vice versa. No reliable correspondence was found between the orientation responses from day 1 and scene identity responses from day 2 (p=.756).

Given that there was a strong correspondence between the orientation responses for both days, a variable called correspondence (*M*=.75, *SD*=.432) was created to examine how much overlap there was between the orientation responses from both days, which indicated the robustness of memory for scene orientation over a delay of one day. Correspondence was 75%, which was consistent with the limited literature on the robustness of orientation-related information in the case still images (Joubert & Oliva, 2010). Performance for the identification of the correct scene was close to ceiling (*M*=.94, *SD*=.23), indicating that long-term memory for scene identity was independent from orientation accuracy and was highly robust over a delay of one day.

Corresponding responses (frequency)		Orientation (day1)		Scene identity (day2)	
		Correct	Incorrect	Correct	Incorrect
Orientation (day2)	Correct	**185**	31	**208**	8
	Incorrect	40	**32**	64	**8**
	Total	216	72	272	16

Table 3. Cross tabulation of frequency of correct and incorrect responses for orientation and scene identity across two days of the Experiment 1B. The above table shows the frequency of correspondence between correct and incorrect orientation responses from day2 according to responses for scene identity from day 2 and orientation from day 1.

Analyses for the independent variables showed that there was a significant effect of the presence of the establishing shot on the first day on the orientation memory (χ^2 (1) = 7.86, p=.005) on the second day (Figure 16).

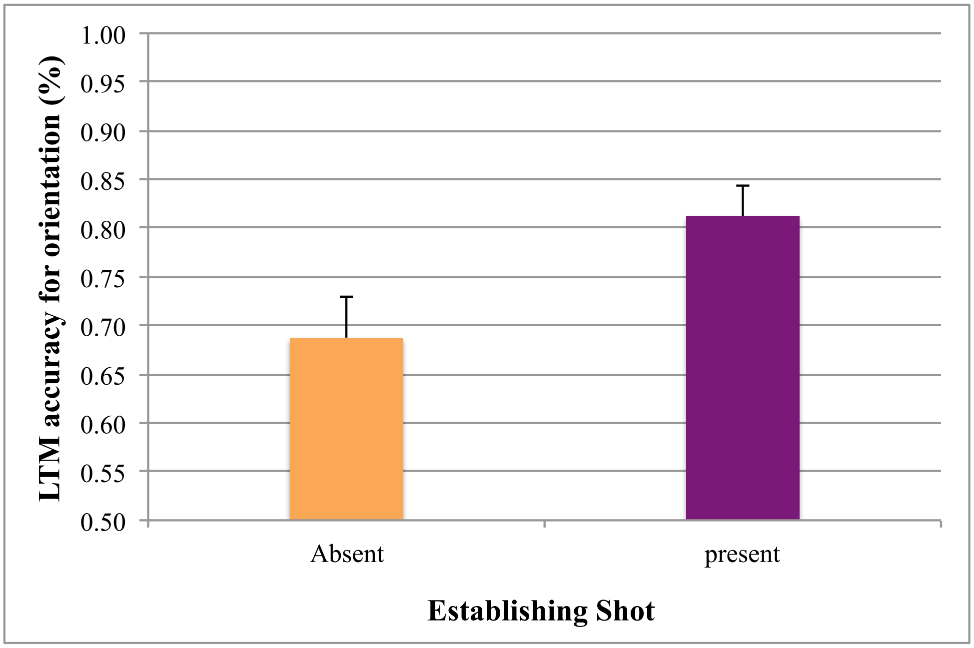


Figure 16. The effect of the establishing shot on orientation memory. The above graph shows the orientation memory on day 2 based on exposure to the establishing shot the first day. Error bars represent one standard error of the mean.

Discussion

Overall, retention of orientation-related information was lower compared to scene identity over the course of a one-day delay. Also, orientation memory was more robust if the subject was exposed to the establishing shot while watching the clips the first day. These results also tell us about the robustness and flexibility of learned spatial maps in visual narratives. In that regard, the correspondence between the two days' responses suggests that people might be sticking with the spatial maps (the relative positions of the actors) they have formed the first day after watching the clips.

It is hard to know for sure how much the orientation memory from the first day versus the spatial relations in the entire sequences plays a role on which orientation was selected the next day. Nonetheless, it is also important to note that subjects were not exposed to the images of establishing shots by themselves as one would see in LTM experiments, which use a series of independent still images. In the case of visual narratives, the consecutive shots in a scene provide a context and strengthen the spatial relations observed in the establishing shot. This may in fact be one of the reason behind the strong correspondence between the two days. This suggests that having more information about the spatial relations in a scene (contextual significance) might have enhanced long-term memory for orientation.

Also, the results showed that the gist of the scene has been well preserved over the course of a one-day delay, regardless of whether the subject was right about the orientation. This suggests that even though information pertaining to orientation might be more open to decay and interference, scene identity appears to be more robust. The performance which was close to ceiling on scene identity suggest that gist of the scene was well preserved and viewers were not confused by alternative scenes from the same movie which had similar characteristics with respect to setting and actors. Further research may illuminate whether this holds true for dynamic stimuli presented as a distractor from the same movie to see whether those could be distinguished as well. Also, individual shots can be used instead of still images of establishing shots to further examine the role of the dynamic nature of the experience on later memory for scene orientation.

Experiment 1C

As the above experiments used mirror images as test stimuli to inquire about the correct position of actors in movie scenes, this experiment is conducted to examine whether there is any inherent preference or sensitivity for subjects with respect to screen orientation (right vs. left). The motivation was to get a baseline measure of preference for image orientation (if any) and to make sure that the mirror images are perceived as natural as the original orientation of movie scenes. Here, “natural” means that they are aesthetically similar and do not feature any odd spatial relations.

Methods

Stimuli and Procedure. In Experiment 1C, participants performed a preference task to indicate any inherent bias for a certain orientation for each test image used in Experiment 1B to consider a potential preference for the original and alternative scenes. In this experiment, participants were asked: "Considering the above images, which orientation appears more natural to you?" They performed a total of 24 trials, which showed the image pairs for the establishing shots for the original and alternative scenes for each movie. This time, instead of a forced-choice paradigm, the participants were given the chance to choose neither of the images, indicating a *neutral* decision. They were given explicit instructions for what the neutral choice represented. They were told to choose that option if they did not have any preference for neither image, meaning that both orientation looked equally natural to them.

Results and Discussion

No preference for orientation was found for movies (*p*=.154) (Table 4) or for scene type (original vs. alternative) (*p*=.351) (Table 5). For non-neutral responses, the frequency of selection for the original images was nearly equal to the mirror images (51.79 % preference for original images vs. 48.21 % preference for mirror images). Also, the images were coded with respect to the number of people in the scene as well as whether the establishing shot was shot from and 180-degree angle (straight shot) versus whether it was slightly at an angle. Those both did not any effects on the preference for orientation. Lastly, the images were also coded for whether the scene was shot form the right versus the left side in cases where a scene was shot at an angle. There was no apparent preference for the side of shooting that affected the orientation preference either (*p*=.543).

		Selection Frequency			
		mirror	**neutral**	**original**	**Total**
Movie name	Erin Brockovich	42	16	54	112
	Five Easy Pieces	43	25	44	112
	Ordinary People	40	29	43	112
	Social Network	40	29	43	112
	Valentine's day	47	17	48	112
	What Women Want	44	25	43	112
Total		256	141	275	672

Table 4. Response frequency with respect to movie in Experiment 2C. The overall frequencies for selecting the original vs. mirror orientation or the neutral option with respect to movies. N=28.

		Selection Frequency			
Scene type		**mirror**	**neutral**	**original**	**Total**
	alternative	142	71	123	336
	original	114	70	152	336
Total		256	141	275	672

Table 5. Response frequency with respect to scene type in Experiment 2C. The overall frequencies for selecting the original vs. mirror orientation or the neutral option with respect to scene type (original vs. alternative). N=28.

The individual shots for the movies used for stimuli in Experiments 1A through 1C were also coded for the frequency of the side of the camera placement (right vs. left) in shot reverse shot (SRS) and over the shoulder (OTS) sequences in the entire movie. Three movies (*Erin Brockovich, Ordinary People, Valentine's Day*) out of 6 were coded to give a representative sample for year of production. The purpose of this coding was to inquire whether there was an inherent preference in the director's choice for portrayal of scenes from one direction versus the other. This could provide more insight into whether viewers are exposed to certain camera angles more often via director influence. If they are sensitive to watch scenes from a specific angle, this could skew their decisions about which orientation looks more natural to them due to mere exposure.

No reliable difference was found in either movie to indicate a preference for camera placement to favor one direction over the other in either the SRS or the OTS sequences. While placing the camera on the left side of the axis of action was more frequent in OTS sequences, the difference was not reliable and this trend was non-existent considering all SRS sequences. This led to the conclusion that at least in our sample, no apparent bias was present for camera placement that would favor choosing one orientation over the other in test trials. We need to analyze a broader number of movies to talk about the existence or non-existence of a trend for the preference of camera placement in movies in general. However, this sample does not support a bias with respect to the directors' choice that could potentially influence the viewers' orientation judgments due to familiarity and exposure.

CHAPTER 4

EXPERIMENT 2: HOW AGENT CUES AFFECT PREDICTIONS OF FUTURE POSITION IN MOVIE SHOTS

The above-discussed studies considered the cases in which all the agents were stationary and did not change their relative positions in a movie scene. As the results indicated, adherence to the 180-degree rule and the presence of an establishing shot affected people's judgments about spatial relations in movie sequences. Movies not only employ rules based on screen position and camera angle but also use agent cues such as gaze and movement direction to signal actor positions. Filmmakers use eyeline match and POV editing to maintain gaze continuity across shots (Bordwell & Thompson, 2003; Chandler, 2009; Mercado, 2011). Also, previous research showed that gaze, posture, head turns and other conversational cues were highly effective in synchronizing viewers' attention to the relevant parts of the screen (Smith, 2012). Especially, eyeline match shots tend be taken from a closer distance and those don't provide much contextual cues for the position of the actors with respect to the environment and each other. As much as actor positions on the screen are congruent, gaze and movement direction cues are consistent for a certain actor, actors display opposite direction cues with respect to each other, when they are conversing. This following study looks more closely into the effect of congruency of agent-related cues in people's predictions about the subsequent position of an actor in a visual narrative.

Not enough is known about the role of agent-related cues (body position, gaze direction etc.) in determining spatial consistency in movies. Research conducted with populations that were naïve to watching movies were successful in understanding certain editing conventions such as conversational turns as long as the setting and the line of action were familiar (Schwan & Ildirar, 2010). Levin and Wang (2009) also showed that people were good at detecting the projected position of an object based on the gaze direction of an actor and viewers were even successful at detecting the location with accuracy even when the order of shots that show the gaze of the actor and the placement of an object were reversed. Similarly, Hymel, Levin and Baker (2016) reported that viewers were at chance (53%) for detecting one instance of order manipulation in an action sequence when they were not distracted with another concurrent task. With respect to agency, research showed that people relied more on the action of an agent rather than his gaze to report the location of an object in a visual scene, especially when those conflicted (Furnaletto et al., 2013). Researchers suggested that participants took the perspective of the agent more often to understand the intention behind their actions (Lozano, Hard & Tversky, 2007; Tversky & Hard, 2009).

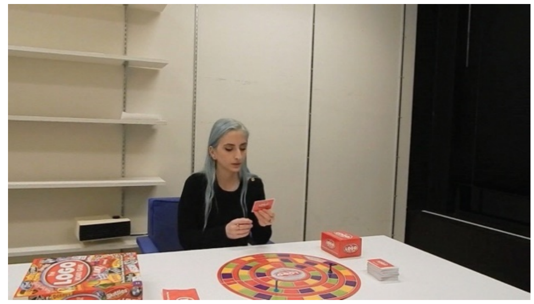
For spatial judgments about an agent, researchers differentiated between two different processes in visual perspective taking: strategies based on line of sight and metal transformation based on agent's perspective (Baker, Levin, & Saylor, 2016; Michelon & Zacks, 2006) Michelon and Zacks (2006) showed that people took less time to respond to questions about the position of an object if the question was about visibility of that object with respect to an avatar and could be answered by relying on the line of sight. This decreased reaction times since the viewer based their decisions on the distance of the object from the avatar on the screen. But, when asked to make a judgment about whether an object is located to the left or right of an avatar, people took more time. This indicated that they engaged in perspective taking and mental transformation of the scene to represent it from the perspective of the avatar.

With these motivations in mind, experiment 2 examined the following: Can viewers correctly identify the upcoming position of an actor in a SRS sequence when the agent-related cues (gaze or body position) were congruent with respect to the actual direction versus screen direction? Does the order of cues affect how well and how fast people decide on the correct position of an actor in a subsequent shot?

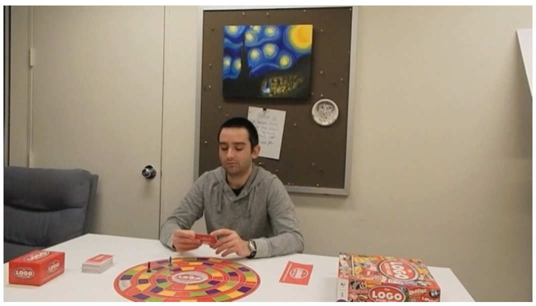
The initial expectation was that people would be performing faster and more accurately when the agent-related cues were congruent, which means that they indicate the same end location in a visual scene, while indicating opposite screen direction. Based on previous research on mental transformations and perspective taking in visual scenes (Baker, Levin, & Saylor, 2016; Michelon & Zacks, 2006), it could be reasonable to expect that if viewers base their decision on perspective taking, they would be taking more time to respond as they would need to make effortful mental transformations. If they use cue direction as a strategy, this would decrease the reaction time by making it easier to predict the upcoming position of an agent. Also, body direction can be expected to have more weight than gaze direction to signal projected position.

Methods

Stimuli. Six activities (playing cards, answering the door, opening a window, making coffee, working on the computer, answering a phone) were filmed in different locations with two sets of actors. Those were in the form of a SRS sequences and comprised of 7 shots, which showed a simple interaction between two actors. In each sequence, only one of the actors gets up from their chair towards a certain direction and the other actor gazes towards a certain direction. The scenes started with an establishing shot, followed by 4 single shots, two for each actor, to familiarize the viewer with the layout and the actor positions with respect to the environment and each other. Those were followed by two shots where the gaze and body direction of actors were manipulated to form 18 variations for each activity. In this experiment, a 3 (gaze side) X 3 (body side) X 2 (order) within-subjects design was used, where *gaze side* and *body side* indicated the direction of the cue (right, left or neutral) and *order* indicated which cue was presented first in the sequence (gaze-first or body-first) (Figure 17). In each clip, the agent cues were either congruent, which means that they indicated the same room direction, and opposite screen direction. Or, they were incongruent, which means that they indicated the opposite room direction but same screen direction. (Please refer to the *Participants* section in Chapter 2 for detailed information on subjects used in each experiment).



a) Shot1- Establishing shot **b) shot 2- single shot actor 1**



c) shot 3- single shot actor 2 **d) shot 6- gaze direction**

e) shot 7- body direction

Figure 17. Examples of movie stimuli used in Experiment 2. The above sequence exemplifies frames taken out of five shots from one of the clips in the experiment condition (gaze-first-congruent) for the board game activity. In this condition, the gaze side cue preceded the body side cue in the sequence and they were consistent with respect to the projected room direction, both pointing out to the same end position for actor 2. However, the cues were incongruent with respect to screen side, meaning that the gaze direction indicated the right side of the screen, while the body side indicated the left side of the screen. Here, shots 4 and 5 were omitted since they displayed a second set of single shots for actors 1 and 2 respectively.

Procedure. In the experiment phase, participants were presented with a set of 18 short clips. Clips were counterbalanced for condition and activity and presented in a random order for each subject. After watching each clip, subjects were asked to perform a 2AFC (two-alternative forced choice) task to choose between two still images that depicted two possible versions of a shot. The images showed still frames from shots, which are "most likely to follow" the sequence they've just watched. In one of the images, the actor who has previously gotten up appeared on the right side of the second actor, while in the other, he or she appeared on the left side of the second actor (Figure 18). The participants were asked: "Which shot is most likely to follow the previous sequence?" RT measures were also taken for each decision.

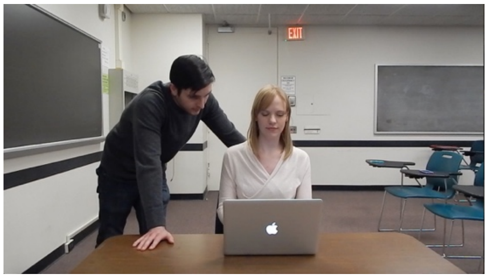
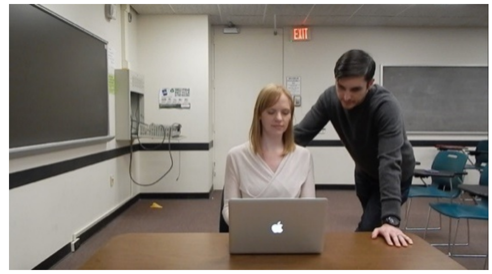
 

Figure 18. Examples of test stimuli used in Experiment 2. The above image pairs show examples of test stimuli for the 2AFC decision task for the predicted position of actor 2. While the above set shows the test stimuli for the board game activity, the middle set shows the window activity and the below set displays the test stimuli for the computer activity.

Results

For analyses, the correct position was considered as the opposite of the screen direction indicated by the actor who got up. Overall, no significant main effects were observed for cue order, gaze and body direction. More interestingly however, there was a significant interaction between gaze and body direction (χ^2 (2, N=42) = 13.39, p <.001) (Figure 19). This indicated that people were better at deciding the correct upcoming position of the actor who gets up, when the gaze and body cues were congruent, both indicating the same room direction. In contrast, their performance suffered when the cues were incongruent, indicating the same screen direction, which presented a conflict with respect to room direction.

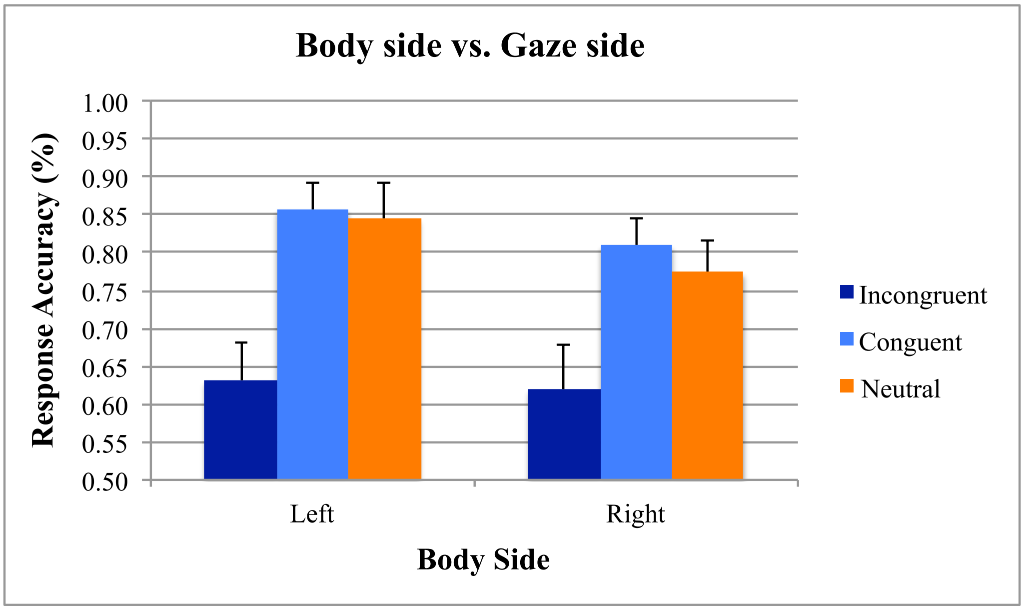


Figure 19. Interaction between gaze side and body side for response accuracy. This figure shows the percent correct for responses related to the projected position of the second actor. Performance was better when the cues were congruent (opposite screen direction) compared to when they were incongruent (same screen direction). Congruent cues resulted in a similar performance to the condition when the gaze was neutral and only the body side indicated a direction. This suggested that people were equally good in considering two cues compared to one, when they were congruent.

Further investigation showed that this result was in fact due to a significant third order interaction between gaze side, body side and cue order (χ^2 (7, N=42) = 21.52, p=.003) (Figure 20). This indicated that the interaction between the gaze and body direction was primarily present for the *body-first* condition, where the cue indicating the body direction was presented earlier than the gaze direction in the sequence. This meant that people depended more on cue congruency in cases where the gaze direction was presented last in the sequence while they mainly depended on the body direction (regardless of the gaze direction) when this was presented last in the sequence. This could in part be due to a recency effect for body direction, which could have made it easier for people to compute the subsequent shot when there was no interference from the gaze side. Also, this could be due to the salience of the motion cue that might have been higher in information value. There were no significant effects observed for activity and order of presentation.

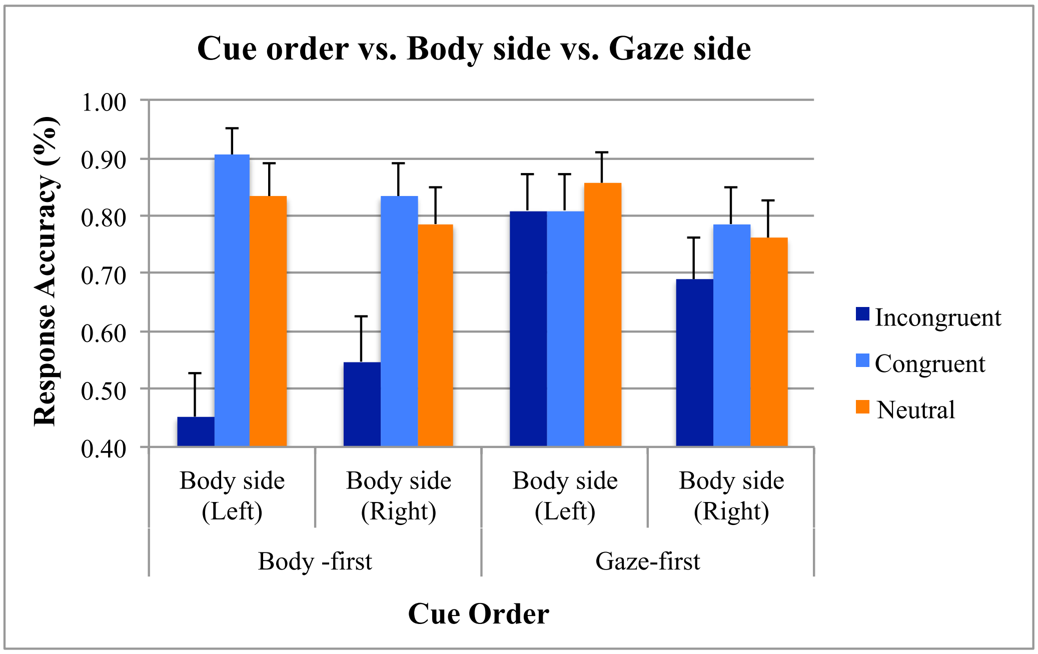


Figure 20. Three-way interaction between gaze side, body side and cue order**.** This figure indicates a significant third-order interaction between cue order, gaze side and body side. While cue congruency was effective in the body-first condition, performance did not depend on cue congruency for the gaze-first condition. People were similar in their decisions for congruent and incongruent videos, when the body direction was presented as the last cue. Error bars represent one standard error of the mean.

Again, there was a significant interaction for gaze and body direction ($F(2,451) = 20.354$, $p<.0001$) for people's reaction times, which indicated that people were slower to make a decision when the cues were not congruent (same screen side), while the reaction times were shorter when gaze side and body side were congruent (relatively consistent with respect to actual position in the scene) (Figure 21).

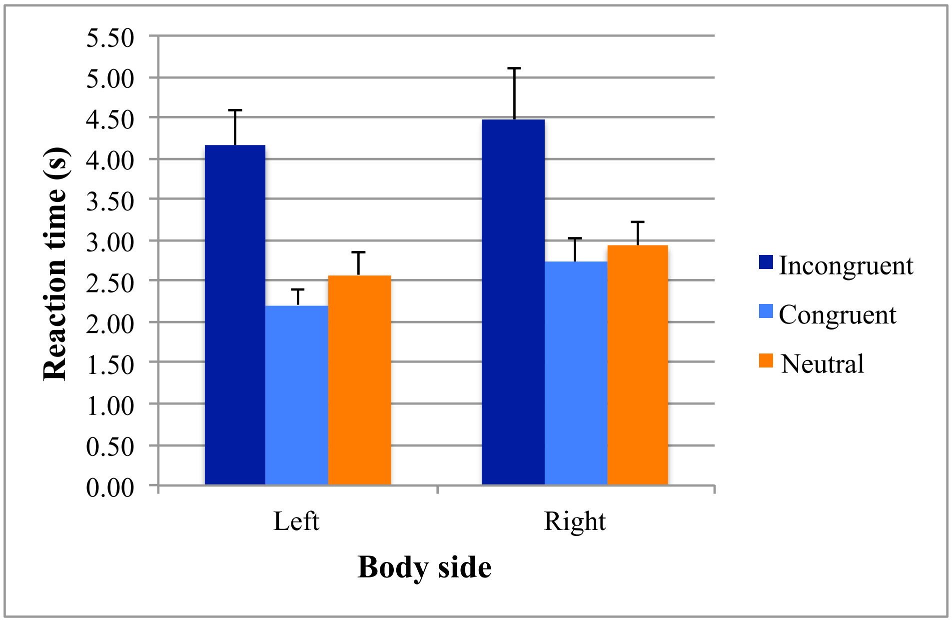


Figure 21. Reaction time results for the interaction between gaze side and body side. The above graph shows mean reaction time measures (in seconds) for different conditions. People were slower to decide the projected position when the cues were incongruent, while the reaction times were shorter when gaze side and body side were congruent. Error bars represent one standard error of the mean.

Discussion

Overall, the results of Experiment 2 showed that congruency of actor cues (gaze and body direction) led to more accurate and faster decision for the projected position of an actor in a movie sequence. This suggests that people used direction cues in the context of the narrative and were good at calculating the projected position based on the direction cues provided in individual shots. This suggests that people are sensitive to relative positions of actors across shots and they can make the mental transformation rather easily when the consecutive shots indicate the same room direction, even when they provide opposite screen direction. We can say that the nature of exposure to spatial information in a visual narrative is different from a static image in the sense that visual narratives provide information about a visual scene from multiple viewpoints in successive shots. In that regard, spatial relations are strengthened through exposure as one becomes more familiar with the relative positions of actors to objects and environmental landmarks, using agent cues to predict projected location might become easier as the viewer has multiple contextual cues available to them to make shot to shot transitions. Further investigation also showed that the congruency effect was primarily present when the body direction came earlier in the sequence than the gaze direction but people relied primarily on body direction when it was presented last in the sequence. This was also the case when it conflicted with the gaze cue. This suggests that when the body direction cue is provided last, it is given more weight possibly because it holds more informative value for the viewer. This is consistent with previous research with still images which also showed that people tend to take the perspective of the agent more often, possibly because it holds more inferential value about the motive of the action. while people were equally good at predicting the correct position of the second actor when the body side cue was presented at the end of the sequence.

Also, people were quicker to decide the projected position of the actor when the cues indicated the same room direction. This again indicated that cue congruency is one of the factors that make it easier for people to make the transitions between shots fast enough to follow the narrative. This is line with research that found that people were faster to react to dynamic scenes involving congruent spatial alignment of objects relative to the environment compared to screen direction (Huff & Schwan, 2012). Also, people took similar time to respond when there was only one cue to consider compared to when there were two but those were congruent. While this seems counterintuitive, it supports the importance of congruency over amount of information, which indicates an efficient usage of editing rules. The overall results of this experiment again support the view that spatial consistency across movie shots appears to be one of the facilitators for viewers to make correct and fast decisions about the spatial relations in subsequent shots. One can say that this exactly what the directors are after when they employ continuity editing rules. Consequently, this experiment shows evidence for the beneficial effects of these practices with respect to efficiency in viewer's reaction times.

CHAPTER 5

EXPERIMENT 3: HOW PERSPECTIVE AFFECTS MEMORY FOR OBJECT POSITIONS IN MOVIE SHOTS

Experiment 3

As discussed above, one essential factor that makes the perception of space in film different than everyday spatial perception is the passive nature of the observer. The following experiment focused on viewers' memory for individual shots in over the shoulder (OTS) sequences. The main motivation in Experiment 3 was to further understand how much detail people register from individual movie shots. More specifically, it investigated whether viewers correctly remembered the viewpoint-related information from a movie shot in an OTS sequence, where two actors' viewpoints are interchangeably depicted.

Studies that examined spatial arrangement of objects in static visual scenes have consistently found viewpoint-dependent representations in static scenes (Shelton & McNamara, 1997). Viewpoint-dependency was also investigated in dynamic scenes, which usually included either single viewpoint presentations without any cuts or the nature of the situation differed from the highly controlled formulaic techniques used in movies. In one such study, Garsoffky et al. (2002) presented viewers with episodes of soccer matches from a certain viewing angle and found that recognition of video stills was better if they were presented from the original viewpoint compared to novel viewpoint. This favored a viewpoint-dependent representation of space in dynamic scenes. They have also found that memory accuracy declined with successive increases in the angle of the distractor scene. Garsoffky et al. (2002) also attended to mimic movies by introducing an angle change halfway into the scene. While this also resulted in viewpoint-dependent recognition, this manipulation didn't capture the complexity of successive angle changes in movies.

More recent studies also showed that semantic content of a dynamic event influenced how much viewers depended on viewpoint (Garsoffky, Huff & Schwan, 2007). In such study, Huff, Schwan and Garsoffky (2011) showed that events, which the authors referred to as "semantically meaningful dynamic scenes" resulted in viewpoint-independent memory of a scene (p. 477). In those cases, the semantic content was abstracted from the specific viewpoint angle. So, there is somewhat mixed evidence about how much viewpoint alignment is necessary to remember a scene correctly. Also, those studies employed simple computer animations with mainly single viewpoint presentations. In contrast, movies offer a richer visual narrative where multiple-view presentations across multiple cuts are presented in a highly structured and regulated fashion.

Experiment 3 focused on location-related information for objects employed by two agents in a SRS sequence. While previous studies suggest that people retain certain information about objects properties across viewpoint changes in movies, position appears to be difficult to retain across viewpoint changes. In a related experiment, Hirose et al. (2010) examined memory for changes in object features across a movie cut and showed that memory for type, color and identity showed a different pattern compared to memory for position. While the former showed a bias towards the post-cut shot and identity and type information was well retained across cuts, information for position was not retained as well as the other properties and did not show the same trend for post-cut dependence. Also, Levin (2010) investigated television series shot from multiple-angles versus sitcoms that were shot from one dominant angle. He showed that viewers found it hard to locate things in the scene if it was shown from multiple angles.

Experiment 3 asked the following specific questions**:** Do viewers remember the correct positions of objects in movie shots according to a specific actor's perspective? Does memory for object location depend on adherence to the 180-degree rule, meaning that, does crossing the 180-degree line adversely affect memory for object locations in movie scenes?

We know from previous research that people are sensitive to identifying 180-degree violations (Kraft et al., 1991; Baker & Levin, 2015). Baker and Levin (2015) argued that due to the limited capacity of visual attention and the adaptive nature of visual perception, people only track the changes in information, which are necessary to follow the narrative. They suggested that spatial relations are only processed when there is a mismatch of information that disrupts the flow. So, according to the authors, spatial consistency violations were one of the ways to activate a comparison between a prior and present representation of an event in working memory. This comparison process has been shown to help memory for information that coincided with event boundaries. People remembered information at event boundaries better than information between boundaries (Swallow et al. 2011). This points to an adaptive way of making sense of visual narratives by processing the most crucial details, which coincide with changes, without wasting cognitive effort on consistent information.

Research also suggested that one of the reasons why people realize these violations could be attributed to sudden changes and inconsistencies associated with the background (Hochberg & Brooks, 1978). Interestingly, Baker and Levin (2015) showed that people's awareness of the 180-degree violation further increased when the background remained the same but only the relative position of the actor has changed. This suggest that the relative location of actors might be more crucial for subjects in detecting screen direction violations because they could be monitoring that information more closely than the background. It can also be argued that having a similar background might have helped detection of the position change by making it more salient. It's important to note that the relative positions of objects with respect to an actor's viewpoint do not change when one crosses the 180-degree line, only the associated background and the relative position of the actor with the respect to the screen changes. So, if those are the main features one depends on to locate objects in movie shots, we should expect a decreased performance in the presence of a 180-degree violation. Otherwise, if one only anchors object position to an actor's perspective, then crossing the axis of action shouldn't affect the accuracy of reports for object positions.

For this experiment, an actual 180-degree violation was employed. For this violation, actual shots that were taken from the other side of the 180-degree line were used for each activity, so that the background also changed with the crossing of the line. While keeping the actor positions consistent across shots, 180-degree rule also keeps the background consistent relative to object positions. Because each shot in an OTS sequence depicts the scene from one of the actor's side, the viewers see the positions of objects from each actor's viewpoint, always associated with the same background.

Methods

Stimuli. Four simple activities were filmed, each as a 5-shot sequence. They depicted an interaction between two actors. All shots were filmed OTS, where the camera was positioned at a 45-degree angle over the shoulder of each actor. The clips were filmed in four different settings with two sets of actors carrying out four *activities* (drinking milk, coffee, soda, and tea). Experiment 3 employed a within-subjects design where the presentation of SRS sequences assumed 2 conditions for scene type: *regular* vs. *violated.* This experiment employed 8 videos where half were conventional SRS sequences where the camera did not cross the 180-degree line (Figure 22). The other half was manipulated so that the 180-degree rule was violated. Those clips started with the same shot from the regular clips, therefore taken from the same angle for the first actor, but crossed the 180-degree line by cutting to the other side of the action line (Figure 23). In these clips, both actors appeared at the same side of the screen. In this experiment, an actual violation of the 180-degree rule was employed with shots that were previously taken from the other side of the 180-degree line for each activity, so the background also changed with the crossing. Clips were counterbalanced for the actor that appeared in the first and last shot and for the side of the screen (right vs. left) where the first actor appeared. (Please refer to the *Participants* section in Chapter 2 for detailed information on subjects used in each experiment).

Figure 22. Example stimuli for the regular condition in Experiment 3. The figure shows example frames from the clips presented to subjects. The above pair of images shows the *milk* activity and depicts a pair of OTS shots. Adhering to the 180-degree rule, each actor appears in a relatively consistent side of the screen across a SRS sequence. The below pair of images shows a pair of OTS shots from the *soda* activity.

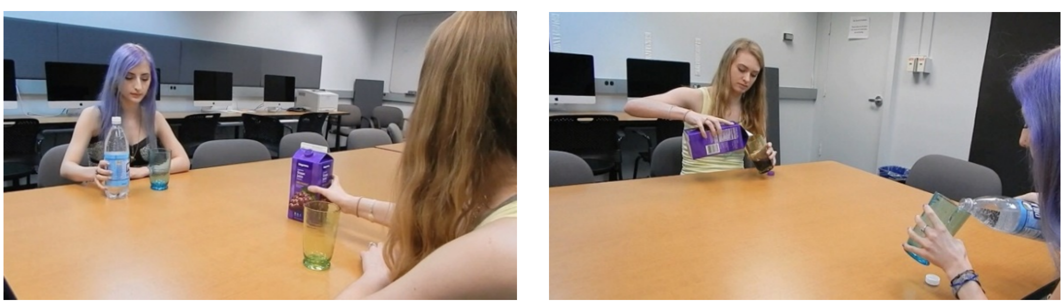


Figure 23. Example stimuli for the violated condition in Experiment 3. This figure shows two still frames from separate shots of the *soda* activity. The left image shows a shot taken from one side of the 180-degree line, while the right image shows the following shot, which violates the 180-degree rule, where the camera crosses the 180-degree line.

Procedure. In the experiment phase, participants watched a total of 8 videos, one for each condition. Activity and condition pairs were counterbalanced between subjects so that each was equally represented. After watching each clip, participants responded to a 2AFC (two-alternative forced choice) task to choose the correct position of objects on the table from each actor's perspective (*test type*) (i.e. "Taking the perspective of the woman with blonde hair, which one of the above images shows the correct position of objects on the table?"). For the test phase, still frames from the shots and their mirror image alternatives were employed (Figure 24). For the first actor, a frame from the third shot was used and for the second actor; a frame from the second shot was used to avoid any primacy or recency effects on memory. In the 2AFC task, the alternative image was the mirror image of the original, this time created by flipping the image 180 degrees horizontally. While this manipulation kept the background the same for each shot, it displayed the objects on the reverse positions compared to the original (i.e. right-left sides). Reaction times were also taken for each trial.

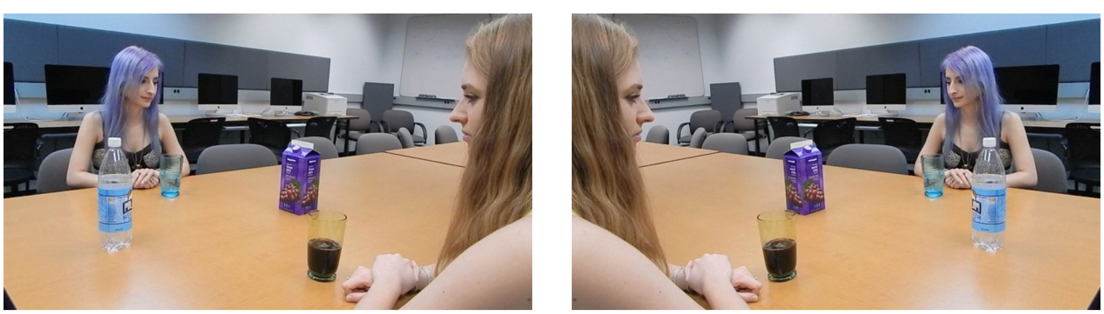


Figure 24. Examples for test stimuli in Experiment 3. This figure shows a pair of images used in one of the test trials to ask the correct location of objects according to the perspective of the woman with blonde hair. While the above left image shows the original frame from the presented clip where the object locations are correct according to reference frame, the above right image shows a mirror alternative where the location of the objects is reversed, therefore incorrect.

Results

Results indicated no reliable effects for the type of sequence (regular vs. violated), test type (1st actor vs. 2nd actor) as well as activity with respect to response accuracy. The recognition accuracy for choosing the image that shows the correct position of objects with respect to each actor's perspective was high in both experimental conditions (regular: *M*=.83 *SEM*=.026; violated: *M*=.86 *SEM*=.030). Interestingly, mean accuracy was higher for scenes where the 180-degree was violated compared to when it was maintained but the difference was not significant. Also, no interaction was found between type of sequence and test type with respect to accuracy (Figure 25). The only effect found on accuracy was that of the reaction time (χ^2 (1, N=28) = 7.99, p=.005), indicating that response latency influenced accuracy.

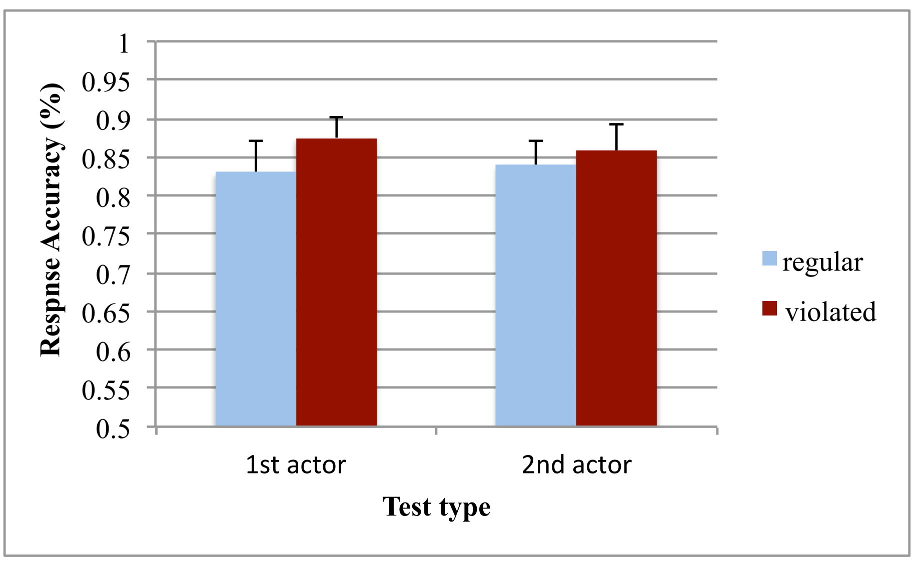


Figure 25. Response accuracy with respect to test type and scene type in Experiment 3. The graph shots that there was no reliable difference between the experimental conditions (regular vs, violated) with respect to response accuracy, and those were not different for each test type (1st actor and 2nd actor). Error bars represent one standard error of the mean.

Looking at the reaction time measures, there was an effect for activity ($F(3,408) = 13.65$; $p < .001$) (Figure 26). Post-hoc analyses with Bonferroni correction indicated that the coffee and milk activities were different from soda and tea which took longer to respond ($p < .001$) and no interaction was found between activity and test type. And while there was a trend for longer reaction times for scenes where the 180-degree was violated, it did not reach significance ($p = .065$).

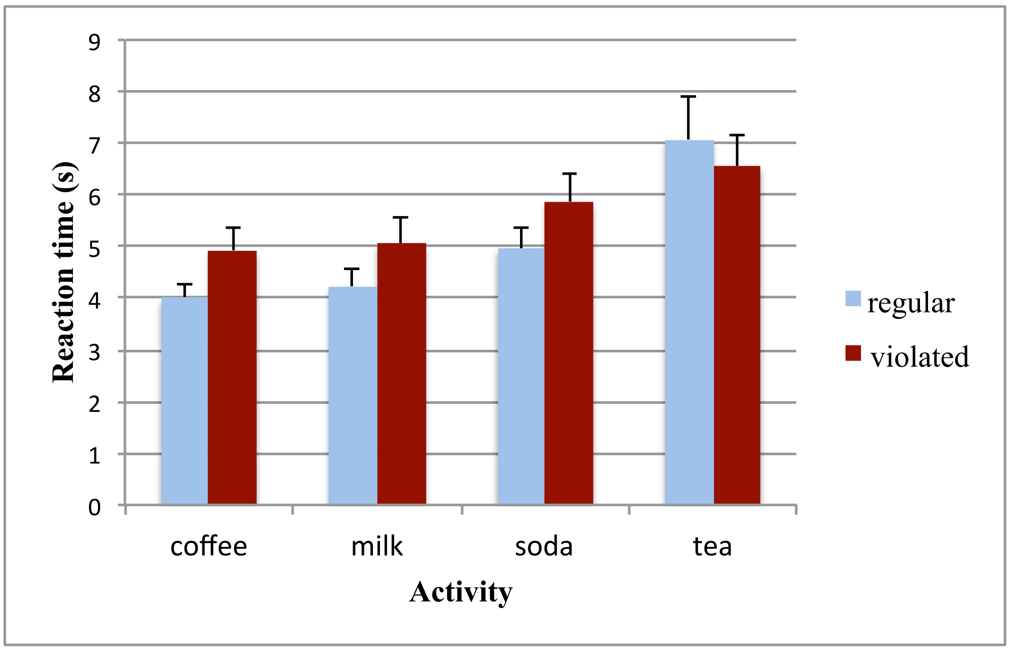


Figure 26. Reaction time results for activity and scene type in Experiment 3. The above graph shows that there was an effect for activity with respect to reaction time results. It took subjects more time to respond to tea and soda activities compared to coffee and milk activities. Also, while the response latencies were longer in the case of violated sequences for most but not all the activities, it did not reach significance and no significant interaction was found for activity and test type. Error bars represent one standard error of the mean.

Discussion

The results indicated that viewers were good at remembering the correct positions of objects according to actors' viewpoints. One reason for that performance could be that in SRS sequences, objects appear on the same relative side of each actor and people view more instances of that relationship across shots. Also, in these activities, the objects were at the center of attention because the actors acted on them, which might have increased their saliency and therefore facilitated memory for position. Research on embodied cognition supports the view that knowledge about action and objects is based on prior active experience and researchers suggested that these formulaic action sequences has conceptual links that facilitate activation of certain action-object pairs easier and faster than others (Borghi, Glenberg, & Kaschak, 2004; Lozano, Hard & Tversky, 2007).

Though, the results should be treated with caution for position related information for peripheral objects or other spatial details that are not at the center of attention. Also, while the overall accuracy was slightly higher in the case of the 180-degree violation, this did not have a reliable effect on performance. This result suggests that the allocated attention to those shots might have increased because of the jarring effect of the violation of the relative position of actors. While this might explain the slightly higher memory for those shots, performance was not significantly different from the non-violated sequences.

The results overall favor the proposition that object positions were anchored to the perspective of the actors and remembered as such even when the actor's screen position was violated when the camera crossed the 180-degree line. While this changed which side of the screen the actor appeared and the background for the actor, the relationship between the positions of the objects with respect to a certain actor's perspective remained consistent across the sequence. If we illustrate this with an example from the soda activity, the grape juice was always on the right and the soda was on the left of the girl with blonde hair while the soda was on the right and grape juice was on the left of the girl with the blue hair. This strengthens the idea that people encode the position of objects with respect to actor's perspectives.

Results also showed an activity effect for reaction times. While people were equally good at identifying the position of objects according to each actor's perspective for all four activities, people responded faster to test stimuli in the coffee and milk activities and they took longer to respond in the soda and tea activities. This might again be due to content and location differences between those activities. Because the background was more similar for actors in the case of soda and tea activities, people might have needed more time to correctly differentiate between the object positions associated with each actor.

Also, a non-significant trend was observed for longer latencies in the case of the violated sequences compared to non-violated sequences. This trend coupled with the activity effect may warrant a further examination of the relationship between 180-degree rule violation and object relations using more activity examples to inquire the potential role of background differences.

CHAPTER 6

EXPERIMENT 4: PERSPECTIVE TAKING FOR VISUAL SCENES INVOLVING TWO AGENTS

In movies, there are no instructions or questions about the object-object or object-agent relationships in a visual scene. We take the understanding of spatial relations at face value based on viewers' comprehension of a coherent narrative. As the above experiments showed, the depiction of a scene from different viewpoints is a challenging task for viewers but they are aided by the editing rules that ensure spatial continuity. While spatial relations between the agents and objects may not often be the primary goal in a movie scene, they need to be congruent in order not to disrupt the narrative flow.

In this chapter, the following set of experiments examined how much perspective taking affects agent-object relations in visual scenes where more than one actor's perspective is depicted. More specifically, they examined which frame of reference viewers adopted to locate objects where more than one agent was present. Tversky and Hard (2009) claimed that a spatial scene does not necessarily have two types of representations, egocentric (with respect to self) or allocentric (with respect to the environment). Being social beings, people can create and maintain multiple representations of space like “other’s perspective”, “my egocentric perspective” and “your egocentric perspective”. Experiment 4 was an adaptation and expansion on a similar paradigm employed in a previous study by Tversky and Hard (2009). The procedure included open-ended questions used in previous related studies (Lozano, Hard, & Tversky, 2007; Tversky & Hard, 2009), this time to get qualitative and more detailed information about how viewers encode spatial relations in movie scenes.

Experiment 4A focused on how the number of visible agents in a movie scene affects which perspective people adopt when describing object relations. Still images depicting movie frames where one or two agents were present were used to describe the relationship between two objects on a table. The motivation was to examine whether people would adopt similar perspectives depending on the number of agents present in the scene. In Experiment 4B, the question was changed to emphasize an action verb. This manipulation is adopted from previous related studies (Lozano, Hard, & Tversky, 2007; Tversky & Hard, 2009). When researchers employed action-based questions, they found that the frequency of taking the perspective of the agent increased, especially when the agent was mentioned in the question. In this experiment, the question only included the action verb without any mention of the agent and or the relationship between objects. One reason for this was not to prime the subjects to only consider one of the agents in scenes where two were present. Also, another reason was to inquire whether putting an emphasis on action only would similarly affect the adoption of the agent's perspective.

In Experiment 4C, a slight alteration in the methodology was implemented. Participants were first introduced to an image showing the establishing shot of the scene before asking about the test image. The purpose of this manipulation was to further examine which frame of reference viewers would adopt if the same scene was first presented from an objective angle as it usually happens in movies. A shot in a movie is not independent of others that precede and follow it. Scenes usually involve establishing shots, which put the actor and object relations in context. The purpose of this manipulation was to see how exposure to an establishing shot would affect people's descriptions with respect to agent-object relations.

With respect to embodied cognition, the relationship between objects in space is based on previous experience in acting on objects. Also, those relationships (i.e. where to put a mug in relation to a coffee pot) are tied to the conceptual knowledge about everyday activities (Carlson & Kenny, 2006). Lozano et al. (2007) suggested that not only our actions, but "perceived" or "simulated" action of another agent also affects perspective taking and people's perception of objects in a visual space related to other agents. Research with static scenes showed that when asked to describe object relations, people tended to take an egocentric perspective more often when the scene did not involve an agent but took the perspective of the agent more frequently when an agent was present in the scene. This tendency increased when people were asked a question that emphasized an action (Furlanetto, Cavallo, Manera, Tversky, & Becchio, 2013; Lozano, Hard, & Tversky, 2007; Tversky & Hard, 2009). Viewers were also more likely to take the perspective of the agent if s/he was interacting with the objects. Also, people relied more on the action cue compared to the gaze cue when those conflicted with each other, which suggested that people considered motion cue to be more informative for the location related information. Taking an agent's perspective is interpreted as having a potential social and adaptive function with respect to predicting people's actions.

Researchers found that the handedness of the viewers also affected whether the agent's perspective is taken when the used hand was varied for the agent when s/he reached for the object (Lozano et al. 2007). In another study, Furnaletto et al. (2013) investigated action and gaze match for how they affect taking the perspective of an agent. They have found that action affected perspective taking more than gaze. Also, interestingly, a mismatch between the agent cues increased perspective taking. In more recent studies, researchers also observed that egocentric judgments about object locations were faster when the asked object is located to the right of the subject and closer to their body (Cavallo, Capozzi, Tversky & Becchio, 2016) More strikingly, according to what the authors have called the *remapping hypothesis*, reaction times were also faster for the opposite direction (left near the avatar) when the participant was asked to take the perspective of the avatar.

While an adaptive and predictive value can be suggested for taking the perspective of an agent, previous studies mainly used single actors in visual scenes where there is no potential conflict of interest or alternative choices. Movies mostly depict more than one actor's viewpoint and encourage involvement in the scene via taking the perspective of the camera. Consequently, an egocentric perspective could be more frequent in SRS sequences where the viewer is encouraged to put himself or herself in the place of the conversational partner. While action of a visible actor may still trigger the potential to take the perspective of that actor and describe the object relations accordingly, how viewers treat such scenes is not very well established. One of the initial expectations for this series of experiments was that scenes where two actors were present would lead to descriptions that include both actors' perspectives. For the question manipulations, how much people put emphasis on the verb versus the mention of agent would be the defining factor in whether similar results from earlier studies would be reached. Also, in terms of the exposure to an establishing shot in Experiment 4C, which depicts the scene from a third person perspective, the frequency of descriptions involving both actors can be expected to increase in scenes where only one actor is visible. In this condition, while the subjects only see one of the actors, they would have an inherent knowledge that the scene also includes the second actor, whose presence is now implied.

Methods

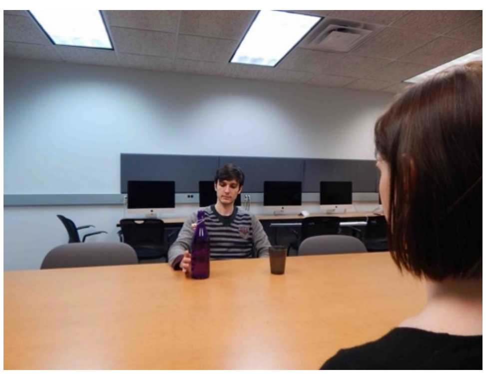
Stimuli. Photographs of two sets of objects (a bottle and a glass) were taken in 4 different settings with 2 pairs of actors. In these photographs, the objects were located next to each other on a table. Photographs were taken from each actor's side for each condition. In this set of experiments: a 3 (contact) X 2 (scene type) between-subjects design was employed. A neutral condition where there was no actor in the scene was also used to get baseline responses. With respect to the *scene type*, the scene either included one actor (*single*), or two actors (*OTS*), where the scene was photographed from an over the shoulder angle of one of the actors. While only one actor was visible in the single condition, both actors were visible in the OTS condition. The *contact* variable had three conditions: *neutral* if the actor did not engage with the objects by looking straight ahead, *gaze* if the actor looked the target object, or *reach* if the actor reached for the target object (Figure 27). (Please refer to the *Participants* section in Chapter 2 for detailed information on subjects used in each experiment).

a) OTS-gaze condition

b) single actor-gaze condition

c) OTS-neutral condition

d) single actor-neutral condition



e) OTS-reach condition

f) single actor-reach condition

Figure 27. Examples of stimuli used in Experiments 4. The above images exemplify the study conditions in one of the settings with respect to the shot type and contact variables.

Procedure. In Experiment 4A, participants were presented with a single photograph of one of the study conditions and responded to the question: “Where is the bottle in relation to the glass?” by typing an open-ended response into text box located below the photograph on the screen.

Experiment 4B used the same procedure as Experiment 4A but this time the question prompt was changed to: “Where is the bottle placed?” instead of "Where is the bottle in relation to the glass?" to emphasize an action verb. They were asked to be as explicit as possible with their answer. The experiment was administered in a between-subjects design and this study was carried out as an addition to other studies, always administered at the beginning. The trials were counterbalanced for setting, actor and the location of the target object.

For Experiment 4C, the subjects were presented with two images consecutively, and the first image showed the establishing shot of the scene taken from a third person perspective, 90-degrees perpendicular to the table. The image showed two people seated at a table facing each other, with two objects (bottle and glass) placed in the middle of them, next to each other (Figure 28). The second image belonged to one of the study conditions used in Experiments 4A and 4B. After subjects were given time to study the establishing shot, they were presented with the test image. Again, this study was administered between subjects where subjects were exposed only to one of the conditions. The question prompt in this case was the same as the Experiment 4A which was: " Where is the bottle in relation to the glass?"

Figure 28. Examples of stimuli used in Experiment 4C. The above figure shows examples for the stimuli used as frames from establishing shots in two of the experiment settings. This view is indented to give an overall third person view of the scene where all the actors and objects were visible but was not portrayed from neither of the actor's point of view.

Results

The answers were initially coded into 5 different categories to get the maximum detail out of the open-ended responses: *self* if the subject responded by taking an egocentric perspective (i.e. “The bottle is located to the left of the glass”), *other* if they responded solely by taking the perspective of the actor facing the camera (i.e. "from the man's perspective", "to the right of the glass"); *self and other* if they adopted both an egocentric perspective and the perspective of the actor (i.e. "The bottle is directly to the left of the class from my perspective, but directly to the right of the class from the man's perspective"); *other and other* if they adopted both actors' perspectives (i.e. “The bottle is to the right of the glass from the perspective of the man and it is to the left of the glass from the perspective of the woman”) or *neutral* if they have not based their description on anyone's perspective but just taken into account environmental landmarks and/or the relationship between the two objects (i.e. "The bottle is about a foot away from the glass"). Two of the categories (self and other, other and other) were then combined into one category called *two perspectives* to better differentiate better from judgments involving a single perspective.

Answers were also coded for *initial response* to reflect which description the subject reported first, with respect to the answers where they have reported more than one perspective and coded for the *mention of self* and *mention of other* in the response (yes vs. no), which indicated whether the subject specifically referred to themselves (i.e. "from my perspective"), or to the actors (i.e. "from the perspective of the guy with the blue hoodie"). The responses were also coded by a research assistant to check for reliability, which was 96% considering all three experiments.

Referring to the data from Experiment 4A, most responses included a frame of reference (94%) even if the question did not ask to report someone's perspective specifically, while only a few people employed a neutral perspective (6%). No effect for scene type was present considering all the response categories. Since there was a visual trend for higher frequency for adopting two perspectives in the OTS conditions compared to egocentric (self) perspective in the single actor condition, a comparison between those responses showed a moderate effect for scene type (χ^2 (1, N=34) =5.44, p =.020) (Figure 29). Further examination also showed that in cases when participants have reported two perspectives, they tended to mention an egocentric perspective first before describing the other in 90% of the cases. No effect was found for contact variable.

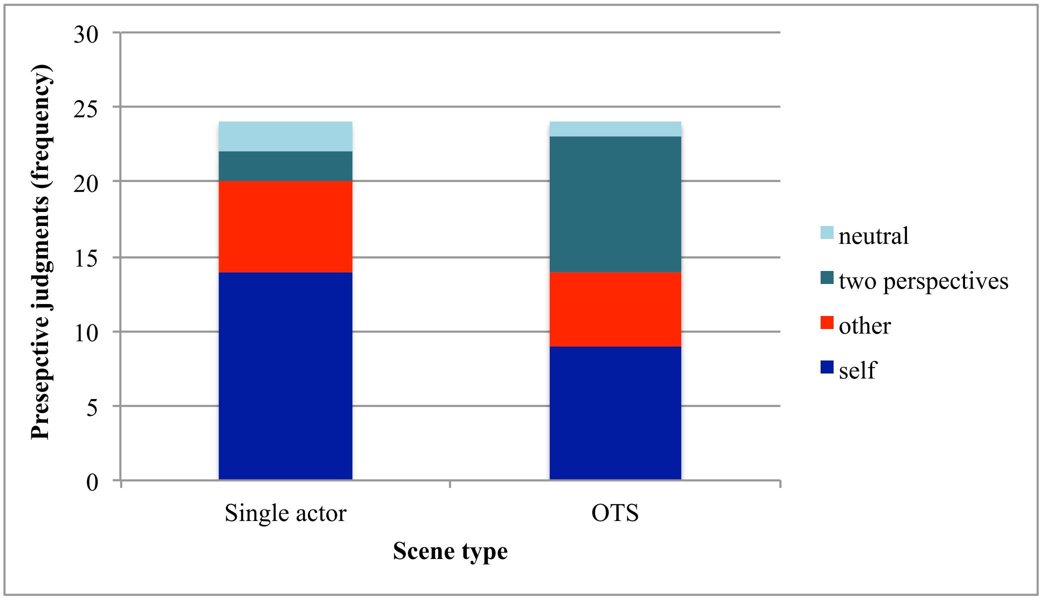


Figure 29. Frequency of perspective judgments with respect to shot type in Experiment 4A. This graph shows the frequency of responses for different decision categories based on the scene type. The question prompt for Experiment 4A was the following: "Where is the bottle in relation to the glass?"

In experiment 4B, which used a placement question to emphasize the action verb, scene type affected the response category (χ^2 (3, N=48) = 14.84, p =.002). People again adopted two perspectives more often in the case of OTS scenes compared to single actor scenes and this time, the adoption of the actor's perspective was more prominent in the single actor condition compared to OTS condition (Figure 30). This effect was consistent with what was found in previous studies that used a similar question prompt. It was interesting to see that not mentioning the agent did not affect the frequency of taking the perspective of the agent. Also, no reliable effect was observed for contact.

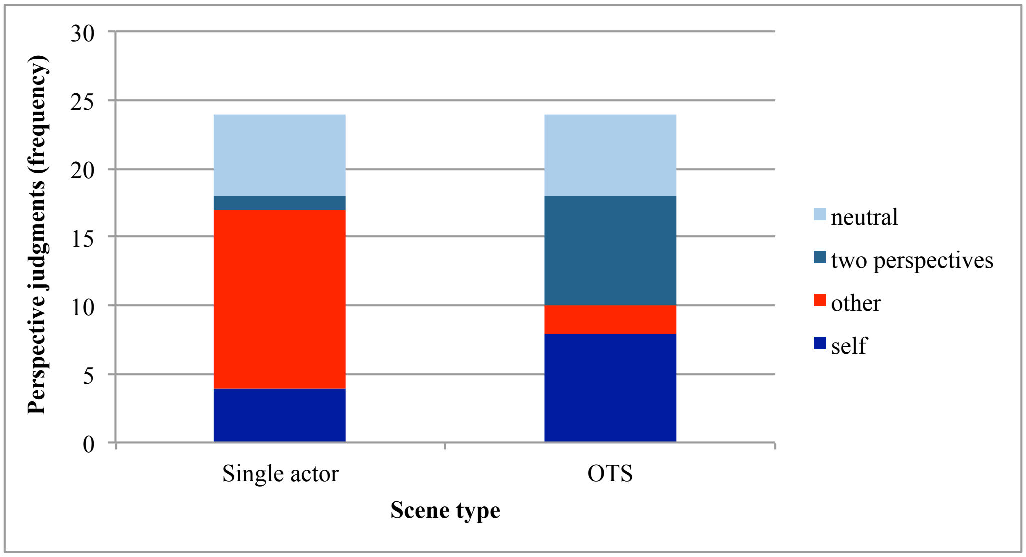


Figure 30. Frequency of perspective judgments with respect to shot type in Experiment 4B. This graph shows the frequency of responses for different decision categories based on the scene type. The question prompt for Experiment 4B was the following: "Where is the bottle placed?"

In Experiment 4C, no effect was observed for both scene type and contact variables. The frequency for egocentric judgments was similar in both conditions. Also, adoption of two perspectives and that of the actor were less frequent in OTS shots compared to single actor shots while neutral judgments were more frequent (Figure 31). Possible implications were discussed when comparing this experiment with Experiment 4A, which used the same stimuli but did not include an establishing shot before the test stimulus.

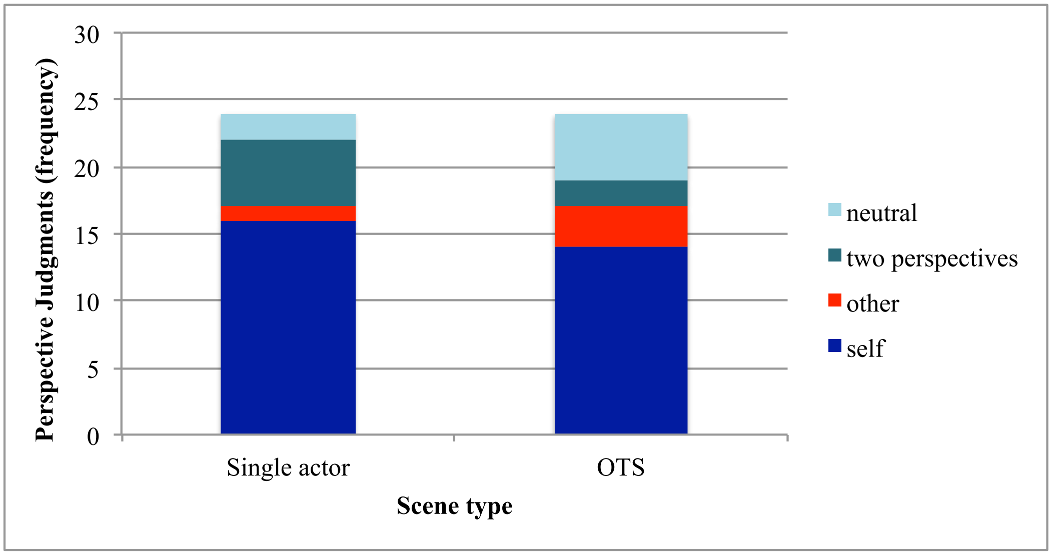


Figure 31. Frequency of perspective judgments with respect to scene type in Experiment 4C. This graph shows the frequency of responses for different decision categories based on the scene type. The experiment 4C included an establishing image introduced before the test image and the question prompt was the following: "Where is the bottle in relation to the glass?"

Secondly, comparisons were made between the experiments 4A and 4B since they used the same stimuli but differed on the question prompt (relationship vs. placement). Comparison was also made between Experiments 4A and 4C as they employed the same question prompt (relationship) but differed on the employment of an establishing shot (present vs. absent) before the test stimuli.

Comparison of experiments 4A and 4B showed effects for scene type (χ^2 (3, N=96) = 15.43, p=.001) and a moderate effect for the question prompt (χ^2 (3, N=96) = 9.67, p=.022) (Figure 32). With respect to scene type, egocentric responses were equally frequent in OTS vs. single actor conditions, people employed both perspectives more frequently in OTS scenes compared to single actor scenes. With respect to the question prompt, egocentric judgments were less frequent in the case of a placement question, which led to more descriptions adopting agent's or neutral perspective. The increase in the frequency of the neutral perspective is especially interesting, suggesting that while an emphasis on relationship led to more perspective taking, emphasis on placement led to less perspective taking, more objective description of a scene. Another difference that was observed between these two experiments was the frequency of mentioning an agent in the description (mention of agent) (χ^2 (1, N=96) = 6.17, p=.013) (Figure 33). While the frequency of mentioning an agent was similar in the case of a relation question, people mentioned the agent more when the question included an action verb.

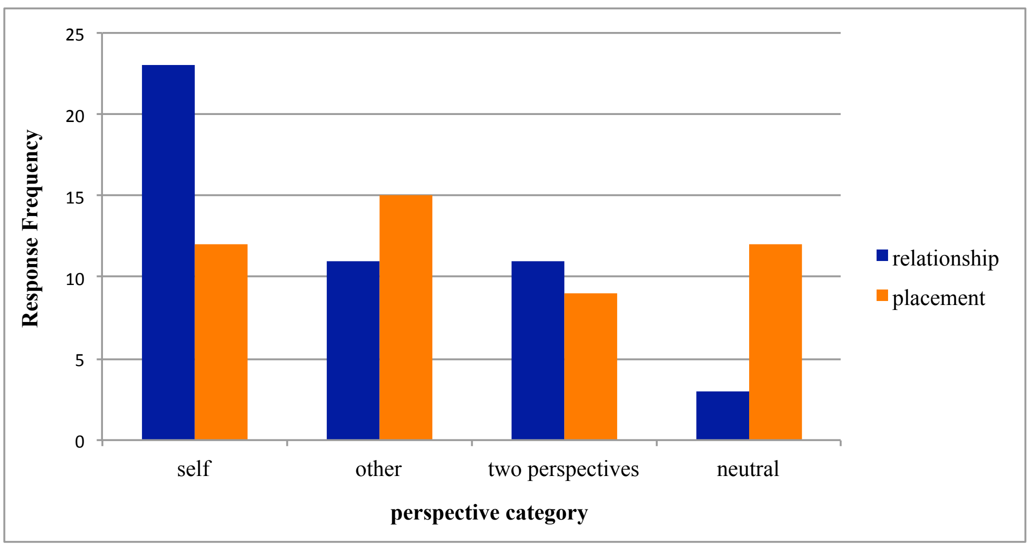


Figure 32. Comparison of perspective judgments between Experiments 4A and 4B. This figure shows the frequency of different perspective judgments between Experiment 4A (relationship) and Experiment 4B (placement).

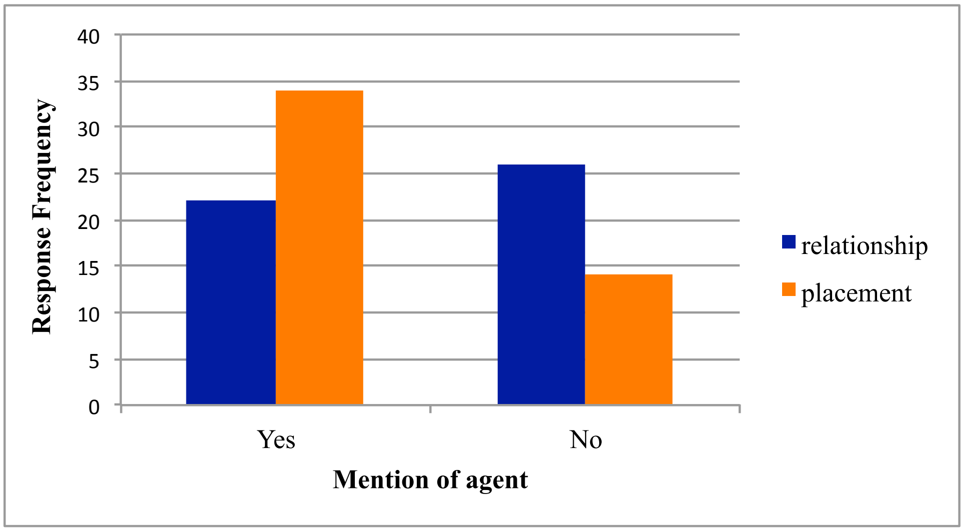


Figure 33. Comparison of the mention of agent between Experiments 4A and 4B. The above figure shows the frequency of mentioning an actor (other) in the response (yes vs. no) for different response categories between Experiment 4A (relationship) and Experiment 4B (placement).

Comparison between experiments 4A and 4C showed a modest effect for the presence of the establishing shot on perspective judgments (χ^2 3, N=96) =8.212, *p* =.042) (Figure 34). While the frequency of responses adopting an egocentric and neutral perspective showed a trend for increased frequency in Experiment 3C, the frequency of adopting two perspectives and the perspective of the agent decreased compared to Experiment 4A. Another difference between the responses in Experiments 4A and 4C was the frequency for including the actor in the description (mention of agent) in the response (χ^2 (1, N=96) = 6.75, *p*=.009) (Figure 35).

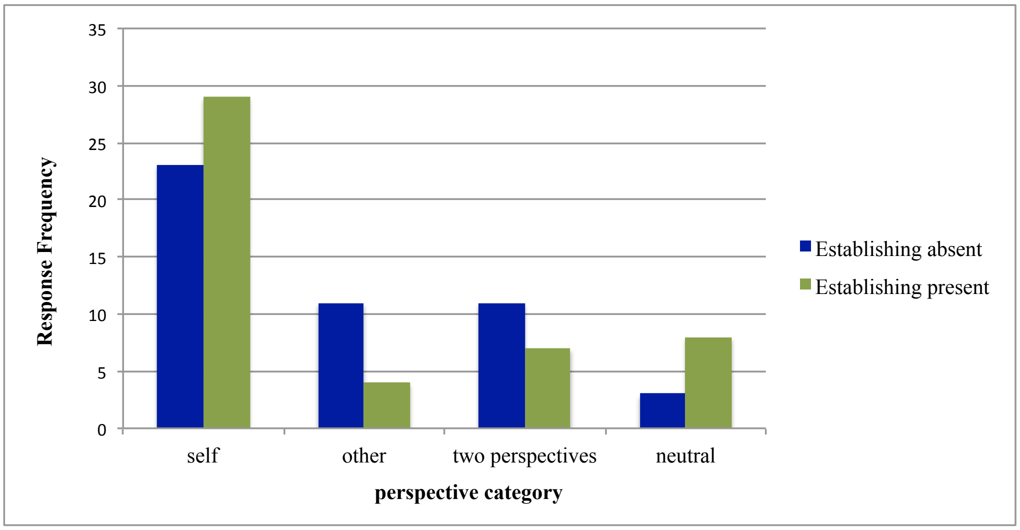


Figure 34. Comparison of perspective judgments between Experiments 4A and 4C. This figure shows the frequency of different perspective judgments between Experiment 4A (no establishing shot) and Experiment 4C (establishing shot).

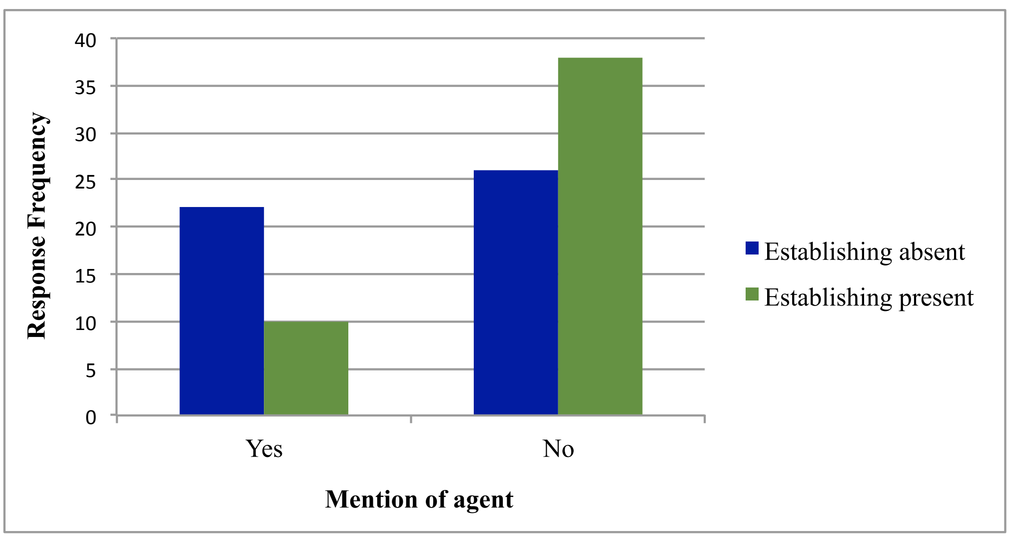


Figure 35. Comparison of the mention of agent between Experiments 4A and 4C. The above figure shows the frequency of mentioning an actor (other) in the response (yes vs. no) for different response categories between Experiment 4A (no establishing shot) and Experiment 4C (establishing shot).

Discussion

Overall, this study provides valuable information by giving a more descriptive account of how people make reference-based assumptions in movie-like scenes. While this study only used static scenes, the results still provide an initial understanding of frame of reference employed with respect to object-agent relations in movies. Comparison between Experiments 4A and 4B showed effects for scene type and question prompt. People adopted the perspectives of both actors more frequently in OTS scenes compared to single actor scenes, where they adopted mostly an egocentric perspective. Also, the presence of an action verb increased the frequency taking the perspective of the agent especially for the single actor scenes. This result is consistent with what was found in previous studies. Also, interestingly, egocentric judgments were less frequent in the case of a placement question, which led to more descriptions adopting the agent's or neutral perspective. The increase in taking the neutral perspective in the placement question was also an interesting result, showing that referring to a relationship was important to elicit perspective taking from viewers. The reliance on only the placement question increased a more objective treatment of the scene

The comparison between Experiment 4A and 4C showed two major differences. When the scene was preceded with an establishing shot, egocentric descriptions increased while other-related descriptions decreased. Also, the referral to the agent in the descriptions decreased. While this was surprising based on initial expectations, it is in line with a narrative approach to perspective taking. We can speculate that this condition more closely approximates what a viewer is exposed to in a movie scene. When subjects were introduced to two consecutive frames, the test image became part of a narrative instead of an independent image. Therefore, subjects might be more likely to report the viewpoint of only one actor, in this case the one whose viewpoint is aligned with their egocentric perspective. This explanation is also consistent with what has been found recently by Cavallo et al. (2016). People were faster to react to viewpoints that aligned with their own perspectives. In the case of movies, that viewpoint is the one of the camera. This suggests that establishing shots, while providing a third-view account of a scene, also familiarize the viewer with the actor-object relations. So, when faced with the depiction of the scene only from one of the character's perspective (in SRS sequences), the viewers might be expected to adopt the viewpoint, which coincides with the camera. This again stresses the passive nature of the viewer being subjected to spatial relations via the changes of the camera. This could be a better use of attentional resources in watching movies, in order to stress the focus on the narrative, which is the directors' primary intention.

CHAPTER 7

CONCLUSIONS

I would like to go back to the same quote from the introduction: “The real world is spatially and temporally continuous: film is not” (Cutting, 2005, p. 9). If so, how do we perceive a subjectively continuous space from a disjointed presentation of successive viewing angles? According to the ecological and evolutionary approach to the perception of film, understanding everyday visual perception and film perception are interconnected (Cutting, 2005, Smith, 2012, Anderson, 1996). Previous research suggested that movies have evolved to fit our perceptual abilities and constraints with respect to attentional patterns and everyday usage of perceptual cues like gaze, posture and gestures (Cutting et al. 2011; Cutting, 2016; Smith, 2012). Cognitive film theory film emphasizes the value of using film as a complex visual stimulus to further the knowledge about our visual perceptual abilities (Bordwell & Thompson, 2003, Bordwell 2002; Levin & Baker, in press). Studying these complex visual stimuli has the potential to teach us more about the mechanisms as well as the potential of our visual system. When we study film, we understand more about how our mind makes sense of these complex visual stimuli and how filmmakers construct movies to accommodate our perceptual tendencies to supply an effortless and enjoyable experience.

In the scope of this dissertation, the experiments discussed here support the suggestion that spatial continuity is one of the core tenants of a visual narrative. It can be said that editing rules that ensure spatial continuity, function as a glue to support the narrative, and give the viewer a foundation for making sense of actor and object relations. In this way, the viewer can allocate the cognitive resources to follow the narrative more effectively. The results indicate that movie techniques use congruency as one the bases for spatial integration of information through the usage of the axis of action (adherence to 180-degree rule across shot sequences) and agent cues (gaze and body direction). Usage of these cues supports a heuristic approach that facilitates the formation of a virtual spatial map out of different viewing angles in the most efficient way, accommodating our attentional tendencies. Their main motivation appears to create a coherent and seamless subjective experience out of otherwise individual shots.

Overall, the conducted experiments showed that people are sensitive to the 180-degree rule as it provides congruent spatial relations between actors. This was also apparent from faster and more consistent spatial judgments for scenes that maintained consistent screen positions for actors. Also, self-reports of viewers indicated that they've realized that the actor positions were reversed when the axis of action was crossed. Reaction times were also affected by number of people present in a movie scene, which was also closely associated with the number of camera positions used. This suggests that people need more time to integrate information into a spatial map when there are more relations to consider. Previous research showed that recent movies have shorter average shot durations and employ more close-ups, especially in conversation scenes (Bordwell, 2002, Cutting & Candan, 2015; Cutting, et al. 2011). In the present movie sample, number of actors in a scene was associated with the varied usage of camera angles, and close-ups were used together with gaze cues and head turns.

One aspect that could be valuable to study further is the use of gaze cues and head turns that were used with close-ups in those situations. While close-ups narrow the focus to one actor and provide fewer cues with respect to the background and the position of actors in relation to others; agent-related cues could be expected to compensate for the lack of spatial information that is otherwise present in wider angle shots. Also, in the present sample, the actors were stationary and did not change positions throughout the scene. Future research focusing on actor and camera movements within a scene to signal actor positions can give us more insight into how people understand the changes from the initial to the final position by making shot transformations.

We also see the value of wider viewing angle in establishing shots, which provide more information about spatial relationships. Establishing shots are also longer in duration (Cutting et al. 2012). These shots appear to have a complimentary yet important role in the representation of space in movies. The results of the experiments discussed here indicate that people can rely on the congruency of screen position in the absence of an establishing shot to form spatial maps. While this indicates that people can integrate information given in separate shots, they are still better at those judgments in the presence of an establishing shot and they depend more on the establishing shot when the directional continuity is not maintained. So, establishing shots have functional value. A similar suggestion can be made based on the observation that spatial judgments were affected by shot duration. In that regard, viewers appear to need a reminder of the overall spatial relations in a visual scene if a certain time has elapsed after they were exposed to a wider-angle view. This also relates to the motivation for using re-establishing shots, which are used to either remind the viewers of the existing relations after elapsed time or to redefine the relative positions of actors. This happens when a new actor or an action line is introduced or when existing actors move to change their positions in the scene. Further research in this area can illuminate when and under which conditions these shots are needed. Also, this can help us understand whether there is a threshold for effective integration of closer angle shots before the employment of those becomes potentially dysfunctional as far as cognitive resources are concerned.

In addition, the experiments showed that viewers were sensitive to the congruency of the agent cues, mainly gaze and body direction. An efficiency argument can also be made based on the observation that viewers reacted faster and made more accurate judgments for the projected position of an agent, when those cues signaled the same room direction. This suggests that people can pay attention to and integrate multiple cues from different agents across movie shots. While this seems a difficult task, directional congruency appears to be a mechanism used to facilitate this task. Also, results showed a heavier reliance on action cues (body direction) compared to gaze direction. It can be suggested that viewers were using their resources efficiently by relying on the most informative cue.

The discussed experiments also suggested that viewers tend to take actors' perspectives into account. Results indicate that this is the case when encoding spatial relations of objects in scenes. The experiments with static images that simulated different camera angles (single shot vs. OTS), also suggest that people make more egocentric judgments to refer to object relations in visual scenes that feature a single actor compared to scenes where two actors are visible and the camera is placed over the shoulder of one actor. In addition, if a scene is first presented from a third person perspective in the case of an establishing shot, where none of the actors' viewpoint is featured, people tend to employ more egocentric judgments. This suggests that representing a scene from a third person viewpoint can be an efficient mechanism with respect to cognitive resources, possibly because it is the viewpoint that aligns with the egocentric viewpoint of the viewer. In this way, each angle is treated as an egocentric angle, the one that aligns with the viewer's perspective. One can expect that this facilitates the integration of shots and makes the transformation between shots easier, maximizing the usage of cognitive resources.

While taking the perspective of the actors is encouraged, movies mostly offer an objective perspective, the one of the camera. The film scholar David Bordwell also stated "traditional film theory ... creates a perspectival eye for cinema, one we call the invisible observer" (1985, p. 9). We can say that the viewer is a silent observer, one that takes the perspective of the camera. While this perspective almost never emulates the exact perspective of an actor, it still encourages the viewer to be more involved in the scene, as the camera angle changes in successive shots to depict each actor's point of view. In this respect, the viewer is passively presented with the space without actively changing their location. The effective nature of the usage of the editing rules that enhance spatial continuity suggest that movies are successful in their attempt to create a continuous simulated space for the viewer.

One of the movie scenes from the present sample can be used to illustrate the main results of this thesis more effectively. In the dinner scene from the movie *Five Easy Pieces* (1970) (scene number 2, refer to Table 1 for the characteristics), multiple actors were present where OTS and single shots were used and close-up shots were accompanied with gaze and body direction cues. First, I'd like to give a brief back story to situate this scene in the movie. *Five Easy Pieces* (1970) is a drama, that tells the story of Robert "Bobby" Dupea (Jack Nicholson), a rebellious pianist from an upper-class family of musicians, who leaves home and starts working in an oil field to experience a different lifestyle. He then returns home after his sister Partita (Lois Smith) informs him that his father is ill. Bobby reluctantly brings his clingy girlfriend Rayette (Karen Black), a waitress, along with him but puts her in a motel instead of offering her to stay in the family home. Bobby starts to realize that he missed what he left behind when he sees his siblings' lifestyle and starts to develop feelings for his brother Carl (Ralph Waite)'s fiancée Catherine Van Oost (Susan Anspach), who is also a pianist. Rayette then surprises him by coming to the house when she runs out of money. In that particular scene, Bobby is having dinner with his sister Partita, brother Carl and his fiancée Catherine, his father accompanied by his nurse and Rayette who joins them.

The scene starts with an establishing shot showing them from a wider third person angle and then it cuts to closer-angle shots, which focus mainly on a conversation among Rayette, Carl, Catherine and Bobby (Figure 36).

a) Establishing shot: *Five Easy Pieces* (1970) dinner scene

b) two shot (Carl and Cartherine)

c) two shot (Rayette and Bobby)

d) Single shot (Catherine)

e) Single shot (Rayette)

f) Single shot (Rayette)

g) Single shot (Bobby)

Figure 36. Example shots taken from the dinner scene of the movie *Five Easy Pieces* (1970). **a)** The image shows a still frame from the establishing shot of the dinner scene. The shot displays a wider-angle view of the table and shows Carl, his father and nurse, Partita, Rayette and Bobby starting from the left side. **b)** The image shows a still frame from the two shot that shows Carl (on screen left) and Catherine (on screen right) looking off screen to the right. **c)** The image shows a single frame from the two shot that shows Bobby (on screen right) and Catherine (on screen left) looking off screen to the left. **d)** The image shows a still frame from the single shot that shows Catherine looking off screen to the right, in the direction of Rayette, who was talking in the previous shot. **e)** The image shows a still frame from the single shot that shows Rayette looking off screen to the left, in the direction of Catherine and Carl, conversing with them. **f)** The image shows a still frame from the same single shot that shows Rayette this time turned to look to the right in the direction of Bobby to address him. **g)** The image shows a still frame from the single shot that shows Bobby looking to the left in the direction of Rayette, not very pleased with her behavior.

At the table, Rayette is telling "everybody" that the house is a nice change compared to the motel where she was staying alone and she had to come uninvited because she ran out of money and was miserable there. Carl is very attentive and nice to her and doesn't understand why she had to stay at the motel. Rayette is very chatty, not very bright, unrefined, and she likes to attract attention to herself. She seems oblivious to Bobby's affection for Catherine and annoys him with his inconsiderate questions.

In this scene, the viewer is informed of each character's position at the table and with respect to each other through the usage of the congruent camera angles as well as gaze and body direction cues. In the original (unaltered) version of the scene, camera is positioned in each shot in a way that maintains the directional continuity so that each character is displayed at the same respective sides of the screen. Two shots provide wider angles so that the viewer is exposed to the relative position of two characters together while single shots are closer but provide consistent screen direction and gaze cues to support the same spatial composition. For example, in this scene, Carl is always positioned at the left side of the screen, with Catherine on his left, looking off the screen to the right. In contrast, Bobby is shown to the right side of the screen with Rayette on his right, looking off the screen to the left (Figure 36: images b and c). While adhering to the 180-degree rule maintain the same relative positions for the actors throughout the scene, violating the rule reverses actor positions. Such a violation for example, would put Carl on the right side of the screen and Bobby on the left side of the screen. This could potentially disorient the viewer by challenging the established spatial relations. This was apparent when viewers were slower in their judgments for actor positions when two halves of a scene displayed reverse spatial relations in Experiment 1A and they did not display a reliable preference for either orientation. This suggests that people consider the scene in its entirety and do not rely only the last shot, especially when an establishing shot is absent. Also, viewers used congruency of screen position in the absence of an establishing shot, which suggests that screen direction is an effective cue in providing the necessary information when viewers need to consider multiple spatial relationships.

One can also illustrate how gaze and body direction cues were employed in movies to support spatial relations, again using the same movie scene. The results of the Experiment 2 indicate that viewers are more accurate in their judgments for the projected position of actors when the gaze and body direction cues are congruent, meaning that they predict the same actual end position for an actor. While that experiment was conducted with stimuli shot in the laboratory with simpler activities between two actors, the use of agent cues has similar characteristics in the above discussed scene from the movie *Five Easy Pieces*. In the original (unaltered) scene, we see Carl and Catherine always looking off screen to the left in single and two shots, which are then followed by showing Bobby and Rayette together or Rayette by herself. This signals that these parties are facing each other and they are on the opposite sides of the table. Head and body turns also support the same spatial relations to facilitate the perception of a coherent spatial map. For example, Rayette always looks to the left when she speaks to Carl and Catherine, which signals that they are on the opposite end of the table, facing her. She turns and looks off the screen to the right when the camera cuts to show Bobby. This signals that Bobby is located next to her on the left, which corresponds the right side of the screen (Figure 36: images d, e, f, g). Gaze cues are especially effective in single shots where screen direction and background cues are less informative. This is achieved by the order of shots in which gaze cues are followed by shots that show where the actor is looking at in the next shot. In that respect, the careful juxtaposition of shots with congruent screen direction creates a coherent spatial map, where viewers come to expect where a certain actor will appear in consecutive shots. This is also in line with the attentional theory of cinematic continuity, which proposes that the attention of the viewers is synchronized to the same locations on the screen through the predictive nature of cues related to screen direction, gaze and body movements across cuts (Smith, 2012).

These results overall support an efficiency argument for the strategic usage of camera angles, which display congruent spatial relations. Spatial relations are provided to viewers with each camera angle that emulates an egocentric perspective. This can be suggested to limit the effort from the viewer's part to compute, reassess or make mental transformations to maintain a coherent spatial representation, which is otherwise effortful. This is consistent with a heuristic approach to spatial updating

(Huff & Schwan, 2012), which suggests that people use the relative screen positions of actors to locate them more easily across a cut instead of spatial alignment based on angular distance, which is a more effortful process. Using these highly predictable sequences provides the viewer with correct predictions for the upcoming positions of actors and objects so that they can expect where to direct their attention on the screen and minimize the usage of cognitive resources.

There were also limitations to the experiments discussed in this dissertation as there were observations that warrant potential for future research. One of limitations can be attributed to the experimental design used in Experiments 1 through 3, which employed a within-subjects design. In a within-subjects design, the same subject is exposed to more than one treatment, which means that he or she is exposed to all the experimental manipulations in a repeated fashion. In contrast, in a between-subjects design, different groups of subjects get exposed to different conditions of the independent variable. While a within-subjects design has its advantages, we should also be weary of its disadvantages and be cautious of interpreting the results keeping in mind the potential confounds that might arise due to the usage of this type of a design. At the on hand, using a within-subjects design is efficient in increasing power and decreasing variance due to individual differences. It increases power basically by decreasing the beta error (false negative), which is the probability of not finding an effect when one exists. A within-subjects design increases power mainly by increasing the sample size through the repeated measures provided by each subject. In this design, the sample size increases by getting multiple measures for each subject compared to a between-subjects design, which would necessitate dividing treatments among different groups of subjects, therefore decreasing subject size. Also, when using a within-subjects design, one also decreases potential confounding variables on the independent variable due to differences coming from using different groups in a between-subjects design.

On the other hand, using a within-subjects design has certain drawbacks. One is a potential learning effect due to practice and the other is a decline in performance due to exhaustion and fatigue. Those are basically “carryover” effects, meaning that the exposure to a previous treatment in the experiment might have a potential effect on the performance of the subject on a subsequent treatment. This could in turn create a confound for the observed effects of the manipulated variables. Basically, the improvement or decline in performance might be due to practice or exhaustion respectively, instead of the difference between treatments. It’s important to be cautious of how individual subjects respond to the treatments during the course of an experiment and whether there are any order effects due to the presentation of the trials. In the present experiments, there were no significant effects observed for order of presentation. Also, the error bars for the means were similar in different conditions, meaning that there was not a big variance in how participants responded to different treatments, even though they were exposed to those in different points in time during the experiment. Also, while a same subject was repeatedly exposed to different treatments, the order of those was randomized so that participant did not know which condition to expect in a subsequent trial and he or she was not exposed to the same activity-condition pair twice. Those suggest that the potential problems of the within-subjects design were minimally effective and the disadvantages could be said to be outweighed by its advantages.

For future directions, research focusing on actor and camera movements within a scene can give us more insight into how people reconcile position changes in consecutive shots. Also, the present results do not give us an indication for the fluctuation of attention while watching these dynamic clips. The usage of a more interactive feature such as an eye-tracking measure can enrich and expand on these results. It would be informative to study the timeframe for learning spatial relations in movies by examining different scenes of different lengths. This could give us an understanding of the tradeoff between scene complexity and exposure time. The function of the background is another potential area of research that can benefit from further examination. As background is another spatial cue that keeps spatial relations constant, it’s important to study its role further as a potential contextual facilitator for spatial judgments in visual narratives.

All in all, spatial continuity appears to have a special place in ensuring narrative continuity. As much as movies take advantage of our perceptual skills, they also provide new and efficient ways of making sense of complex visual stimuli. Now that personal technological devices encourage more visual involvement from the part of the viewer in everyday life, understanding how these media use our perceptual abilities to give us a smoother, effortless and enjoyable experience is even more crucial. Consequently, studying the underlying mechanisms of perspective taking in movies can give us more information about how and why these attract viewers and how much the positioning of the camera can affect the involvement of the viewer. As we understand more about how we perceive and understand these complex visual stimuli, we get to understand more about the limits and constraints of our visual system and its potentials.

As much as real life offers a spatiotemporal continuity that the film does not, it is still very much accepted by our visual system, which has the potential to adapt to new ways of visual experience. We can speculate that watching movies equips the viewer with strategies to become efficient viewers in time, via heuristics like relative screen position. In a way, viewers learn a new language, for which they already have the building blocks. As we know from previous research, people from communities that were never exposed to movies could still make sense of various movie techniques when the story and the setting were familiar (Schwan & Ildirar, 2010). While this suggests a reliance on universal perceptual abilities in designing movies, certain techniques (especially point of view editing) were still harder for novices to decipher and required some exposure and familiarization with movies.

Research indicates that movies have evolved to better fit to our perceptual abilities and preferences (Cutting et al. 2011; Cutting, 2016; Smith, 2012). On the one hand, filmmakers study and understand which techniques are more captivating to the audience and were better received in the past through movies' success. On the other hand, viewers learn what to expect and how to make sense of these techniques, which in turn make them more sophisticated viewers. In this way, they become more experienced with exposure and can adapt to the complexities of visual media, which is integrated into our daily lives more and more each day. Eventually, movies have become better in exploiting as well as fitting to viewers' attentional preferences so that viewers would be more captivated and motivated to watch a movie. One way to ensure that is to make the transitions between shots as seamless as possible. The experiments in this dissertation support this kind of an efficiency hypothesis, specifically for spatial information presented in movies. They suggest that viewers are better and faster in their spatial judgments for shot combinations that conform to the editing techniques, which ensure and facilitate a continuous narrative.

REFERENCES

Allen, G. C., Siegel, A. W., & Rosinski, R. R. (1978). The role of perceptual context in structuring spatial knowledge. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(6), 617–630, doi: 10.1037/0278-7393.4.6.617.

Anderson, J. (1996). *The Reality of Illusion: An Ecological Approach to Cognitive Film Theory*. Carbondale, IL: Southern Illinois University Press.

Angelone, B. L., Levin, D. T., & Simons, D. J. (2003). The relationship between change detection and recognition of centrally attended objects in motion pictures. *Perception*, *32*(8), 947–962, doi:10.1068/p5079.

Baker, L. & Levin, D. T. (2015). The role of relational triggers in event perception. *Cognition*, *136*, 14-29, doi: 10.1016/j.cognition.2014.11.030.

Baker, L. J., Levin, D. T., & Saylor, M. M. (2016). The extent of default visual perspective taking in complex layouts. *Journal of Experimental Psychology: Human Perception and Performance*. *42*(4), 508-16, doi: 10.1037/xhp0000164.

Berliner, T., & Cohen, D. J. (2011). The illusion of continuity: Active perception and the classical editing system. *Journal of Film and Video*, *63*(1), 44-63.

Bordwell, D. (1985). *Narration in the fiction film*. Madison, WI: The University of Wisconsin Press.

Bordwell, D. (2002). Intensified continuity visual style in contemporary American film. *Film Quarterly*, *55*(3), 16-28, doi: 10.1525/fq.2002.55.3.16.

Bordwell, D. & Thompson, K. (2003). *Film art: An introduction*. New York: McGraw-Hill.

Borghi, A. M., Glenberg, A. M., & Kaschak, M. P. (2004). Putting words in perspective. *Memory & Cognition*, *32*(6), 863-873, doi:10.3758/BF03196865.

Brady, T.F., Konkle, T., Alvarez, G.A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325–14329, doi: 10.1073_pnas.0803390105.

Carlson, L., & Kenny, R. (2006). Interpreting spatial terms involves simulating interactions. *Psychonomic Bulletin & Review*, *13*(4), 682-688, doi:10.3758/BF03193981.

Cavallo, A., Capozzi,F, Tversky, B., & Becchio, C. (2016). When far becomes near: perspective taking induces social remapping of spatial situations. *Psychological Science, 28*(1), 69-79, doi: 10.1177/0956797616672464.

Chandler, G. (2009). *Film editing*. Studio City, CA: Michael Wiese Productions.

Cumming, S., Greenberg, G., & Kelly, R. (2017). Conventions of viewpoint coherence in Film. *Philosophers Imprint*, *17*(1), 1-28, permanent URL: http://hdl.handle.net/2027/spo.3521354.0017.001.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71, doi:10.1006/cogp.1998.0681.

Cutting, J. E. (2005). Perceiving scenes in film and in the world. In J. D. Anderson & B. F. Anderson (Eds.) *Moving image theory: Ecological considerations* (pp. 9-17). Carbondale, IL: University of Southern Illinois Press.

Cutting, J. E. (2016). The evolution of pace in popular movies. *Cognitive Research: Principles and Implications*, *1*(1), 30, doi:10.1186/s41235-016-0029-0.

Cutting, J. E., DeLong, J. E., & Nothelfer, C. E. (2010). Attention and the evolution of Hollywood film. *Psychological Science*, *21*(3), 432-439, doi:10.1177/0956797610361679.

Cutting, J. E., Brunick, K. L., DeLong, J. E., Iricinschi, C., & Candan, A. (2011). Quicker, faster, darker: Changes in Hollywood film over 75 years. *i-Perception*, *2*(6), 569-576, doi:10.1068/i0441aap.

Cutting, J. E., Brunick, K. L., & Candan, A. (2012). Perceiving event dynamics and parsing Hollywood films. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(6), 1476-1490, doi:10.1037/a0027737.

Cutting, J. E., & Candan, A. (2015). Shot durations, shot classes, and the increased pace of popular movies. *Projections*, *9*(2), 40-62, doi:10:3167/proj.2015.090204.

Diwadkar, V. A., & McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science*, *8*(4), 302-307, doi: 10.1111/j.1467-9280.1997.tb00442.x.

Franklin, N., Tversky, B., & Coon, V. (1992). Switching points of view in spatial mental models. *Memory & Cognition*, *20*(5), 507-518, doi:10.3758/BF03199583.

Furlanetto T, Cavallo A, Manera V, Tversky B, & Becchio C. (2013). Through your eyes: Incongruence of gaze and action increases spontaneous perspective taking. *Frontiers in Human Neuroscience*, *7*, 82-86, doi:10.3389/fnhum.2013.00455.

Garsoffky, B., Schwan, S., & Hesse, F. W. (2002). Viewpoint dependency in recognition of dynamic scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1035-1050, doi: 10.1037//0278-7393.28.6.1035.

Garsoffky, B., Huff, M., & Schwan, S. (2007). Changing viewpoints during dynamic events. *Perception*, *36*(3), 366-374, doi: 10.1068/p5645.

Greene, M. R. & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, *58*(2), 137-179, doi:10.1016/j.cogpsych.2008.06.001.

Goldzieher, M. J., Andrews, S., & Harris, I. M. (2017). Two scenes or not two scenes: The effects of stimulus repetition and view-similarity on scene categorization from brief displays. *Memory & Cognition*, *45*(1), 49-62, doi: 10.3758/s13421-016-0640-9.

Hard, B. M., Recchia, G., and Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology: General.140*(4), 586-604, doi: 10.1037/a0024310.

Hirose, Y., Kennedy, A., & Tatler, B. W. (2010). Perception and memory across viewpoint changes in moving images. *Journal of Vision*, *10*(4), 1-19, doi: 10.1167/10.4.2.

Hochberg, J., & Brooks, V. (1978). The perception of motion pictures. In E. C. Carterette & M. P. Friedman (Eds.). *Handbook of perception* (Vol. X, pp. 259–304). New York: Academic Press.

Huff, M., Jahn, G., & Schwan, S. (2009). Tracking multiple objects across abrupt viewpoint changes, *Visual Cognition*, *17*(3), 297-306, doi:10.1080/13506280802061838.

Huff, M., Schwan, S., & Garsoffky, B. (2011). When movement patterns turn into events: Implications for the recognition of spatial configurations from different viewpoints. *Journal of Cognitive Psychology*, *23*(4), 476-484, doi: 10.1080/20445911.2011.541152.

Huff, M., & Schwan, S. (2012). Do not cross the line: Heuristic spatial updating in dynamic scenes. *Psychonomic Bulletin & Review*, *19*(6), 1065-1072, doi: 10.3758/s13423-012-0293-z.

Hymel, A., Levin, D.T., & Baker, L.J. (2016). Default processing of event sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(2), doi:235-246. 10.1037/xhp0000082.

Ildirar, S. & Schwan, S. (2014). First-time viewers' comprehension of films: Bridging shot transitions. *British Journal of Psychology*, *106*(1), 133-151, doi: 10.1111/bjop.12069.

Johnston, M. B., & Hayes, A. (2000). An experimental comparison of viewpoint-specific and viewpoint-independent models of object representation. *The Quarterly Journal of Experimental Psychology: Section A*, *53*(3), 792-824, doi: 10.1080/713755903.

Jiang, Y. V., & Swallow, K. M. (2013). Spatial reference frame of incidentally learned attention. *Cognition*, *126*(3), 378–390, doi: 10.1016/j.cognition.2012.10.011.

Jiang, Y. V., Swallow, K. M., & Capistrano, C. G. (2013). Visual search and location probability learning from variable perspectives. *Journal of Vision*, *13*(6), 1-13, doi: 10.1167/13.6.13.

Germeys, F., & d'Ydewalle, G. (2007). The psychology of film: perceiving beyond the cut. *Psychological Research*, *71*(4), 458-466, doi: 10.1007/s00426-005-0025-3.

Joubert, O. R. Oliva, A. (2010). Similar scenes seen: What are the limits of the visual long-term memory fidelity? [Abstract]. *Journal of Vision*, 10(7):745, doi:10.1167/10.7.745.

Kraft, R. N., Cantor, P., & Gottdiener, C. (1991). The coherence of visual narratives. *Communication Research*, *18*(5), 601-616, doi:10.1177/009365091018005002.

Konkle,T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed thank you think: The role of categories in visual long-term memory. *Psychological Science*, *21*(11), 1551-1556, doi: 10.1177/0956797610385359.

Kwok, S. C. & Macaluso, E. (2015). Immediate memory for "when, where, and what": Short-delay retrieval using dynamic naturalistic material. *Human Brain Mapping*, *36*(7), 2495-2513, doi: 10.1002/hbm.22787.

Lawson, R., & Humphreys, G. W. (1996). View specificity in object processing: Evidence from picture matching. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(2), 395–416, doi: 10.1037/0096-1523.22.2.395.

Levin, D. T. & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin Review*, *4*(4), 501–506, doi:10.3758/BF03214339.

Levin, D. T., & Simons, D. J. (2000). Perceiving stability in a changing world: combining shots and integrating views in motion pictures and the real world. *Media Psychology*, *2*(4), 357-380, doi: 10.1207/S1532785XMEP0204_03.

Levin, D. T., & Saylor, M. M. (2008). Shining spotlights, zooming lenses, grabbing hands, and pecking chickens: The ebb and flow of attention during events. In T. Shipley & J. Zacks (Eds.), *Understanding events: From perception to action* (pp. 522-554). New York: Oxford University Press, doi: 10.1093/acprof:oso/9780195188370.003.0022.

Levin, D. T. & Wang, C. (2009). Spatial representations in cognitive science and film. *Projections*, *3*(1), 24-52, doi: 10:3167/proj.2009.030103.

Levin. D.T. (2010). Spatial representations of the sets of familiar and unfamiliar television programs. *Media Psychology 13*(1), 54-76, doi: 10.1080/15213260903563006.

Levin, D. T., & Baker, L. J. (in press). Bridging views in cinema: A review of the art and science of view integration. *WIREs Cognitive Science*, doi: 10.1002/wcs.1436.

Lozano, S. C., Hard, B. M., & Tversky, B. (2007). Putting action in perspective. *Cognition*, *103*(3), 480-490, doi:10.1016/j.cognition.2006.04.010.

Magliano, J.P., Dijkstra, K., & Zwaan, R.A. (1996). Generating predictive inferences while viewing a movie. *Discourse Processes, 22*(3), 199–224, doi: 10.1080/01638539609544973.

Magliano, J. P., Miller, J., & Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology*, *15*(5), 533-545, doi:10.10002/acp.724.

Magliano, J. P., & Zacks, J. M. (2011). The impact of continuity editing in narrative film on event segmentation. *Cognitive Science*, *35*(8), 1489-1517, doi:10.1111/j.1551-6709.2011.01202.x.

Mascelli, J. V. (1965). *The five C's of cinematography*. Los Angeles: Silman-James Press.

Matthews, W. J., Benjamin, C., & Osborne, C. (2007). Memory for moving and static images. *Psychonomic Bulletin & Review*, *14*(5), 989–993, doi:10.3758/BF03194133.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge, UK: Cambridge University Press.

Mercado, G. (2011). *The filmmaker's eye: Learning (and breaking) the rules of cinematic composition*. Burlington, MA: Focal Press.

Meyerhoff, H. S., Huff, M., Papenmeier, F., Jahn, G., & Schwan, S. (2011). Continuous visual cues trigger automatic spatial target updating in dynamic scenes. *Cognition*, *121*(1), 73-82, doi: 10.1016/j.cognition.2011.06.001.

Meyerhoff, H. S., & Huff, M. (2016). Semantic congruency but not temporal synchrony enhances long-term memory performance for audio-visual scenes. *Memory & Cognition*, *44*(3), 390-402, doi:10.3758/s13421-015-0575-6.

Messaris, P. (1994). *Visual literacy: Image, mind, and reality*. Boulder, CO: *WestviewPress*.

Michelon, P. & Zacks, J. M. (2006). Two kinds of visual perspective-taking. *Perception and Psychophysics*, *68*(2), 327-337, doi:10.3758/BF03193680.

Mou, W., McNamara, T. P., Valiquette, C. M., & Rump, B. (2004). Allocentric and egocentric updating of spatial memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 142, doi: 10.1037/0278-7393.30.1.142.

Oliva, A. & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, *155*, 23-36, doi:10.1016/S0079-6123(06)55002-2.

Schacter, D. L., Norman, K. A. & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology*, *49*(1), 289–318, doi:10.1146/annurev.psych. 49.1.289.

Schwan, S. & Ildirar, S. (2010). Watching film for the first time: How adult viewers interpret perceptual discontinuities in film. *Psychological Science*, *21*(7), 970-976, doi: 10.177/0956797610372632.

Shelton, A. L & McNamara, T. P. (2004). Spatial memory and perspective taking. *Memory & Cognition*, *32*(3), 416-426, doi:10.3758/BF03195835.

Simons, D. J., & Wang, R. F. (1998). Perceiving real-world viewpoint changes. *Psychological Science*, *9*(4), 315–320, doi: 10.1111/1467-9280.00062.

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes during a real-world interaction. *Psychonomic Bulletin & Review*, *5*(4), 644–649, doi:10.3758/BF03208840.

Simons, D. J. (2000). Current approaches to change blindness. *Visual cognition*, *7*(1-3), 1-15, doi: 10.1080/135062800394658.

Smith, T. J., & Henderson, J. M. (2008). Edit blindness: The relationship between attention and global change blindness in dynamic scenes. *Journal of Eye Movement Research*, *2*(2):6, 1-17, doi:10.16910/jemr.2.2.6.

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

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

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, *25*(2), 207–222, doi: 10.1080/14640747308400340.

Swallow, K. M., Barch, D. M., Head, D., Maley, C. J., Holder, D. & Zacks, J. M. (2011). Changes in events alter how people remember recent information. *Journal of Cognitive Neuroscience*, *23*(5), 1052-1064, doi:10.1162/jocn.2010.21524.

Tatler, B. W., & Land, M. F. (2011). Vision and the representation of the surroundings in spatial memory. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *366*(1564), 596–610, doi:10.1098/rstb.2010.0188.

Tversky, B. (2004). Narratives of space, time, and life. *Mind & Language*, 19(4), 380-392, doi: 10.1111/j.0268-1064.2004.00264.x.

Tversky, B. & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective taking. *Cognition*, *110*(1), 124-129, doi:10.1016/j.cognition.2008.10.008.

Wang, R. F., & Spelke, E. S. (2000). Updating egocentric representations in human navigation. *Cognition*, *77*(3), 215–250, doi: 10.1016/S0010-0277(00)00105-0.

Wang, R. F., Crowell, J. A., Simons, D. J., Irwin, D. E., Kramer, A. F., Ambinder, M. S., et al. (2006). Spatial updating relies on an egocentric representation of space: Effects of the number of objects. *Psychonomic Bulletin & Review*, *13*(2), 281–286, doi:10.3758/BF03193844.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychological Bulletin*, *133*(2), 273, doi: 10.1037/0033-2909.133.2.273.

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

Zacks, J. M., Speer, N. K., Swallow, K. M., & Maley, C. J. (2010). The brain's cutting- room floor: Segmentation of narrative cinema. *Frontiers in Human Neuroscience*, *4*, 1-15, doi: 10.3389/fnhum.2010.00168.

Zacks, J. M., Kurby, C. A., Eisenberg, M. L., & Haroutunian, N. (2011). Prediction error associated with the perceptual segmentation of naturalistic events. *Journal of Cognitive Neuroscience*, *23*(12), 4057-4066, doi: 10.1162/jocn_a_00078.

Zwaan, R. A., Langston, M.C., & Graesser, A.C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science, 6*, 292-297, doi: 10.1111/j.1467-9280.1995.tb00513.x.

Zwaan, R. A. & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162–185, doi: 10.1037/0033-2909.123.2.162.

FILMOGRAPHY

Rafelson, B. (1970). *Five Easy Pieces*. USA: Columbia Pictures Corporation.

Redford, R. (1980). *Ordinary People*. USA: Paramount Pictures.

Soderbergh, S. (2000). *Erin Brockovich*. USA: Universal Pictures.

Meyers, N. (2000). *What Women Want*. USA: Paramount Pictures.

Marshall, G. (2010). *Valentine's Day*. USA: New Line Cinema.

Fincher, D. (2010). *The Social Network.* USA: Columbia Pictures