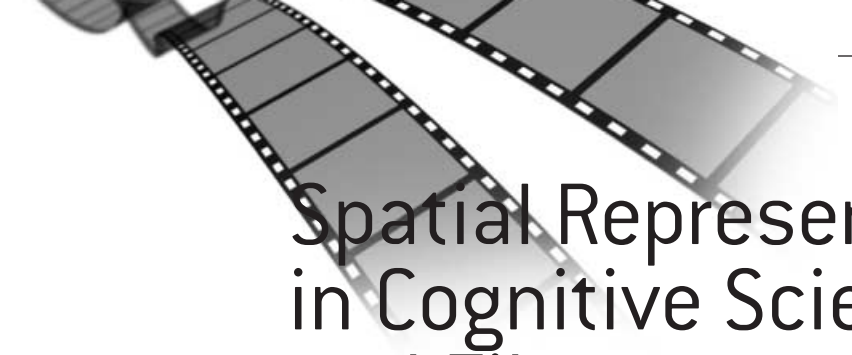


Spatial Representation in Cognitive Science and Film

Daniel T. Levin and Caryn Wang

Abstract: Levin and Simons (2000) argued that perceptual experience in film and the real world share a deep similarity in that both rely on inferences that visual properties are stable across views. This article argues that the perception and representation of visual space also reveal deep commonalities between film and the real world. The article reviews psychological research on visual space that suggests that we not only attend to similar spatial cues both in film and in nonmediated settings, but also that the rules for combining and selecting among these cues are similar. In exploring these links, it becomes clear that there is a bidirectional relationship between cognitive psychology and film editing that allows each to provide important insights about the other.

Keywords: cognitive development, intentionality, representation, spatial cognition, spatial memory

There are several moments during the film *Drugstore Cowboy* (1989) in which we are surprised to see an extreme close-up of some object. In one scene, Bob, the main character, receives a beating from the police. In the midst of a relatively conventional set of reverse angles, we suddenly see an extreme close-up of the knot of a police officer's tie. In a scene in a hospital, after an abortive attempt to steal drugs, Bob ducks into a bathroom. While he is washing up, we suddenly see a small piece of the PA speaker as it emits a mundane announcement. In a scene later in the film, we see an extreme close-up of the printing on the top of a light bulb as it is turned on. These shots are interesting for a variety of reasons. First, they entreat the audience to ask why they are suddenly looking at this tiny slice of the world, and then perhaps to realize that the answer must have something to do with the internal mental states of the main character. These close-ups probably have a lot to do Bob's slightly offkilter and obsessive drug-fueled paranoia. Thus the audience is forced to use visual awareness in an unusual and interesting way.

These shots are subtle variants of the jarring discontinuities characteristic of new-wave cinema because they disrupt the typically seamless process of

Volume 3, Issue 1, Summer 2009: 24–52 © Berghahn Journals
ISSN 1934-9688 (Print), ISSN 1934-9696 (Online)

relating the visual world to the internal thoughts that make it coherent. This kind of disruption can bring to the fore two fundamental issues of visual cognition. First, in forcing the audience to focus on one idiosyncratic piece of the visual world, these insert shots make clear the degree to which our awareness of specific visual objects and properties is highly selective. This kind of selectivity has been the focus of much research documenting the surprising degree to which we are aware of far fewer visual properties than one might think. For example, a phenomenon referred to as "change blindness" occurs when people fail to detect visual changes that occur across views (i.e., continuity errors; Levin and Simons 1997), even when they are attending to the changing object (see Simons and Levin 1997; Rensink 2002). This implies that people fail to create mental representations of most visual properties, or that these representations are not effectively tracked across views (see Simons 2000).

Because the links between the awareness of visual properties and film have been explored previously (Levin and Simons 2000), this article focuses on a second fundamental question of visual perception: the understanding of *where* something is rather than *what* it is. In some of the insert shots in *Drugstore Cowboy* it is fairly clear where the depicted objects are; for example, viewers know where the knot of a tie is on a person's body and the beating sequence makes it fairly clear that the tie is being worn by the officer who is hitting Bob. In other cases, however, it is not at all clear where the object is. It is extraordinarily difficult to know where the PA speaker is because its location is specified only in a close-up of Bob's reflection in a mirror, as he briefly glances at the speaker *after* it has been shown. We showed this clip to a group of colleagues, and asked them the location of the speaker. They were baffled after their first viewing, but after several more viewings, they came to some agreement that the speaker was near the door to the restroom or in the hall outside. The close-up of the light bulb switching on occurs as Bob enters his dark apartment and hits a wall switch. Although it is fairly clear that the bulb is in the sole lamp in the room and thus must be the one we have seen, this conclusion results from deliberation rather than immediate perception.

These shots, and many others like them, provoke a series of questions about spatial perception and memory that are of interest not only within the context of understanding film, but also at the more basic level of psychological processes. For example, is the location of some shots perceived effortlessly while others require thought? If so, what distinguishes the effortlessly perceived shots from the ones perceived with more difficulty? How does knowledge about the objects in a scene, and more generally about people and events, affect our ability to understand these locations? Is this knowledge applied only with deliberation, or is it used more automatically? Given the different kinds of spatial information available, how flexible are viewers in switching from one kind of information to another? Do viewers get comfort-

able using one kind of information in a given film or scene, leaving them vulnerable to a sudden switch in emphasis? Once people learn where things are in a scene, do they remember this information so that they can more effectively perceive the same setting the next time they view it?

Psychological research documenting how people perceive the locations of things in their environment can help us develop at least partial answers to some of these questions. This research is not only helpful in explaining the basis for well-known editing heuristics, but may also help formalize other less systematized intuitions. However, more interesting from our point of view is the possibility that a careful analysis of film editing practice can provide the basis for new psychological research about perception in general.

This article combines research in visual cognition and examples from films to establish four central hypotheses about the perception of space in film. First, at least some of these processes appear to start with an automatic coding of at least some spatial information. Second, this coding relies on a range of specific cues that can be combined flexibly in a context-sensitive way. Third, these initial spatial encodings are often modified and organized by basic concepts about events and the behavior and thoughts of sentient agents. Finally, online episodic encodings of the spatial layout of scenes can be built into longer-term durable representations of complex spaces, but only if certain kinds of information are consistent between these episodic representations.

The 180-Degree Rule and Basic Research in Spatial Coding

If research documenting change blindness demonstrates that viewers often fail to create representations of objects and object properties, then we might ask whether representations of other kinds of visual information, such as the spatial location of objects, are encoded more readily. This is particularly interesting in the context of this paper because both filmmakers and psychologists have concluded that people often do encode the locations of important objects, even when there is no explicit demand for them to do so. In the case of film, the tendency to "spatialize" important objects appears to underlie the well-known 180-degree rule whereby camera positions are constrained during filming to one side of a line defined between centers of attention (see, e.g., Arijon 1976; Reisz and Miller 1953).

In the typical example of the 180-degree rule, the centers of attention are two characters looking at each other as they converse. As illustrated in Figure 1, a scene might start with one or more establishing shots that show both characters, and then would continue with close-ups of each character as they take turns speaking. If the director defines a line of regard between the two characters, and keeps all camera positions on one side of the line,

Both filmmakers and psychologists have concluded that people often do encode the locations of important objects, even when there is no explicit demand for them to do so.

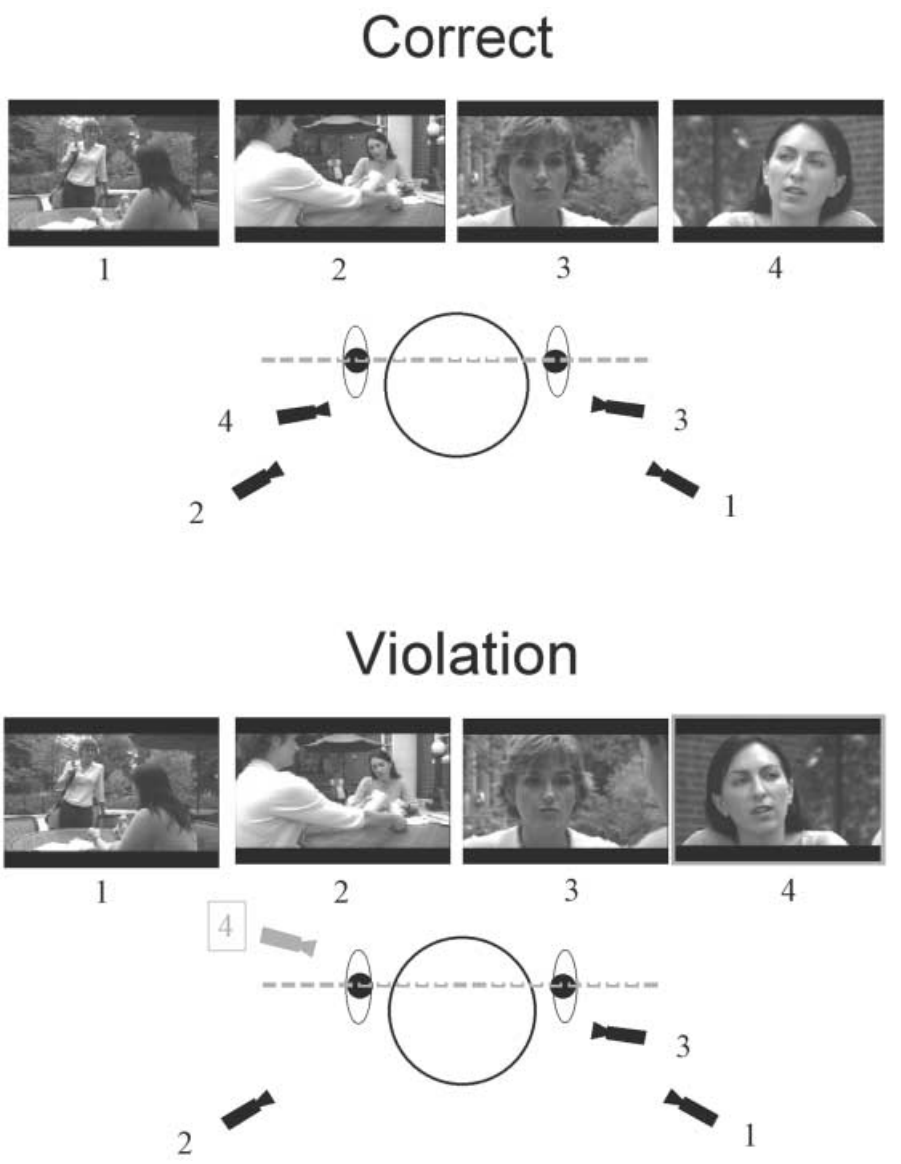


Figure 1. The 180-degree rule. The correct images depict a sequence for four shots that are consistent with the 180-degree rule. The camera positions that produced each shot are shown from a birds-eye view. The violation occurs on the fourth shot: the camera has moved across the eyeline, reversing the established gaze direction of the actor in shot 4 who now looks to the left instead of the right as she had been in the first three shots.

then it will be possible to edit the shots together in seamless continuity because each character will be consistently looking in a corresponding direction off-screen at the other. The idea is that violating this rule will disturb the viewer's sense of space, either making it appear as though an actor is now looking into empty space or possibly that the actor herself has moved. Al-

though no research we know of has tested whether 180-degree violations are noticeable, several experiments have confirmed that violations result in a less accurate memory for the locations of objects in the scene (Frith and Robinson 1975; Kraft 1987), especially where there is no establishing shot (Kraft, Cantor, and Gottdiener 1991).

We start by reviewing some basic research that reveals cognitive principles relied on by the 180-degree rule. It is important to note that we narrow our focus to a subset of spatial coding processes in film. Thus, we focus on spatial coding processes that allow the integration of scene-based information from cuts between different viewpoints (for discussion of other issues of spatial perception that might be relevant to film, see Hochberg 1986; Intraub and Richardson 1989; Sedgwick 1982).

Automatic Coding of Spatial Layout

A range of findings suggests that adults encode at least some kinds of spatial information automatically. That is, they appear to process spatial information whether or not they are explicitly required to, and this processing requires relatively few cognitive resources. It is important to note that it is very difficult to argue that some task is automatic in that it requires *no* cognitive resources, and is triggered *every* time a relevant stimulus is presented. We ask whether spatial memories are generally created with relatively little effort, in a range of situations that extend well beyond tasks that overtly require these representations. In early change detection work, Simons (1996) found that changes to the layout of an array of objects were much more easily detected than changes to object properties. For example, having subjects perform a verbal task while they were trying to detect changes interfered with detection of property changes, but did not interfere with detection of changes to the spatial layout of objects.

In addition to the Simons research, several other experiments seem to demonstrate that spatial representations, of at least small-scale space, are created automatically. In these experiments, participants have been presented with a set of objects arrayed on a table or in a subset of a grid of cubbyholes. Participants are asked to remember the identity of the objects, while the objects' locations are not mentioned. Later, participants are given a surprise test of their recall of the locations of the objects, and, generally, are reasonably successful at remembering them (Mandler, Seegmiller and Day 1977; Schulman 1973; Shoqeirat and Mayes 1991). Thus, it appears as though people sometimes encode objects' locations whether they are told to or not.

In the next section, we describe research exploring the specific cues on which spatial representations rely. In discussing two key cues (gaze direction and spatial geometry), a common theme emerges. Both of these cues seem to be coded automatically, but these initial automatic processes are often sup-

plemented by more sophisticated, nonautomatic ones that allow for the perception of the increased flexibility and durability of spatial representations.

Encoding Gaze: Percepts Leading to Concepts

A key part of research on spatial representations explains filmmakers' practice of organizing views around the gaze direction of actors. Research exploring the phenomenon of joint attention has demonstrated that even young children learn to look at objects others have looked at. Six month-old infants will follow an adult's gaze to one of two objects that both the adult and child can see. When they are twelve months old, infants can follow an adult's gaze to objects that are behind the child (Butterworth and Jarrett 1991). These gaze-following skills are important for a wide variety of tasks. At a deeper level, many researchers have emphasized that gaze is an important "window to the soul" for children: the act of looking at something implies that the looker has internal thoughts that guide their exploration of the world. Therefore, research on attention posits a sequence of developmental stages that starts with a simple gaze that can eventually be understood as indicative of the beliefs, desires, and goals that cause human action (Butler, Caron and Brooks 2000; Moll and Tomasello 2004; also see Woodward 2005). This understanding is usually referred to as a "Theory of Mind," and its roots in the visual perception of spatial gaze have been reviewed repeatedly (Flavell, Green and Flavell 1990; Gopnik, Slaughter and Meltzhoff 1994).

One of the most interesting things about gaze perception and its subsequent elaboration into a full-fledged Theory of Mind is that the child's early processes remain important for later gaze perception and Theory of Mind in adults. This continuity can be seen in several proposed frameworks that posit a distinction between two kinds of processes (Apperly, Riggs, Simpson, Chiavarino, and Dana Samson 2006; Baron-Cohen 1995; Leslie, Friedman, and German 2004). The first process is automatic and includes gaze perception and simple inferences about the mental representations that guide gaze. These inferences include true beliefs and straightforward desires (e.g., Joe looks at the cookie we offer him, and we know he thinks the cookie is available to him and that he wants to eat it). In contrast, other more sophisticated inferences such as predictions based on false beliefs are not automatic and require explicit thought (e.g., we see Joe looking at the cookie jar from which we have taken the last cookie; we know that he wrongly thinks there is still a cookie in the jar, thus we know that his desire for a cookie will be disappointed). These models are interesting because they provide a way of distinguishing easy inferences that require little cognitive support because they are automatic from more complex thoughts that require more support, either in the form of explanation or additional processing time. This dual process model implies that gaze-based spatial organization inherent to the 180-degree rule may rely on

some very basic and general perceptual processes, but that these are closely related to more sophisticated and meaningful cognitions about the actual meaning of events.

Spatial Geometry: Basic Processes and Cognitive Elaboration

Hermer and Spelke (1994) completed a series of experiments exploring how young children orient themselves in space. They were interested in understanding whether the ability to use some kinds of spatial information is fundamental, and could therefore be used even by young children, and whether the ability to use other kinds of information would require the increased cognitive sophistication that comes with age. In these experiments, two-year-old children were taken into a small rectangular room-like space created by sheets hanging from pipes. Once in the room, the children observed an adult hide a toy behind one of the corners of the room. The children where then picked up and spun several times to disorient them. Once they were placed back on the floor, the children were asked to find the toy. The top icon in Figure 2 depicts the mean number of times (out of four) that the children searched in each of the corners of the room, with the (relative) correct location marked by an X. As is clear from the figure, children did not always find the object, but their pattern of errors was revealing: their searches were largely confined to the correct location and the one diagonally opposite it. This implies that the children were using the geometry of the room to locate the object; that is, by searching for it in the locations where the short wall was on their left and the long wall on their right. Because there were no other means of finding the object, it should not be surprising that they searched both corners that satisfied this cue. In fact, not only do adults do the same thing, but previous research had suggested that rats do it as well. This led Hermer and Spelke to hypothesize that the use of this kind of geometric information characterizes a hard-wired brain system that operates in fundamentally the same way across children, adults, and other species.

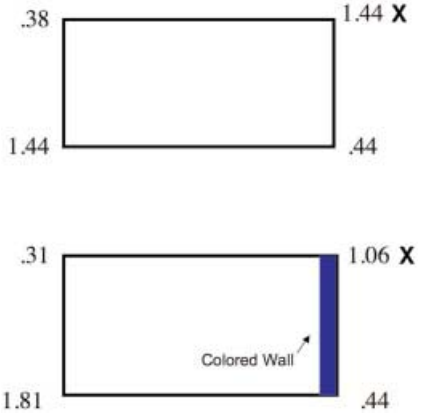


Figure 2. Results from Hermer and Spelke (1994). Number of searches (out of four) from each corner of a rectangular room by disoriented two-year-olds.

Of course, simply observing that everyone uses geometric information in a room where this is the only available information is hardly evidence that there is anything special about this cue. Therefore, Hermer and Spelke did further experiments, adding another salient cue to the room. They replaced one of the short walls with a blue sheet. Now, the target would be behind the wall with the blue sheet or opposite it. Combined with the geometric information that children clearly used in the previous study, this would seem to make for an easy task, allowing for the combined geometric and static property cues to

fully disambiguate the location of the target. Indeed, when adults took part in the study, they had no problem finding the target. However, when three-year-olds participated in the experiment, they did exactly the same thing they had done before: they searched in the two correct diagonally opposite corners, failing to use the blue wall to find the target. It is important to note that it is not as if the children failed to notice the wall; they often spontaneously commented on the enormous blue wall after leaving the room!

These data reveal two important principles. First, some kinds of spatial information are fundamental, and second, using this fundamental information requires relatively little effort. If children and rats can effectively use spatial geometry, then it is likely that the use of spatial geometry does not require much effort to use. Further evidence in support of this hypothesis comes from research using Hermer and Spelke's reorientation task in adults who are, or are not, performing a distracting task while seeing the object as it is hidden. Unsurprisingly, adults who are able to give the target-search task their full attention can use both room geometry and wall color to find the target, and so rarely make errors. In contrast, adults who are asked to do a complex verbal or spatial distraction task sometimes fail to use the blue wall, searching only in both geometrically correct corners (Hermer-Vazquez, Spelke, and Katsnelson 1999; Ratliff and Newcombe 2008). Thus, in these situations, the use of geometry survives distraction, while the use of the blue wall does not.

Adaptive Combination of Spatial Cues

Subsequent research has demonstrated that there are situations where children and distracted adults can successfully use nongeometric cues to find hidden objects (see Lee, Shusterman and Spelke 2006; Newcombe and Sluzenski 2004; Ratliff and Newcombe 2008). It therefore appears as though geometric information is important, but that it is one of several important cues, alongside gaze and other stable nongeometric visual properties and landmarks. Findings such as these have led researchers to hypothesize that there may be fundamentally different kinds of spatial information (and processes), but that people can use a wide variety of cognitive skills to combine them. A full accounting of what kinds of spatial information people use requires careful consideration not only of these skills, but also of the relative salience of each source of information in a given setting. According to the recently proposed adaptive combination view (Ratliff and Newcombe 2008), adults strategically select from a range of geometric and nongeometric spatial cues, based on both short-term and long-term learning about what kinds of information are generally useful and reliable in a given setting. For example, one might learn to inhibit the use of geometric cues when navigating the long hallways in a large building because the geometry of hallways is too complex to be use-

ful in this setting, and instead rely on nongeometric landmarks such as elevators or wall color.

Combined with research exploring the effects of distraction on spatial representation, the adaptive combination view provides an interesting starting point for understanding film editing, but even more interesting is the possibility that film editing can provide insight into how this model might be extended. In the former instance, psychological research has explored a set of cues and delineated general principles that suggest they are combined flexibly, all of which is potentially helpful in understanding how film editing conventions help audiences represent visual scenes. Similar to research exploring Theory of Mind, a key conclusion about these data is that some kinds of fundamental and/or salient spatial information can be used with minimal effort, while more subtle information requires focused attention and elaboration to be used effectively. However, current psychological theory has less to say about how one might select one cue over the other in any given circumstance. This is where a close analysis of film editing practice could reveal insights about the specific balance of spatial cues in a variety of important settings. The next section explores the 180-degree rule and discusses how continuity editing may reveal just such a balance.

The 180-degree Rule and Its Exceptions

One of our favorite 180-degree violations appears in the film *Twins* (1988). The violation occurs during a conversation between the enormous Arnold Schwarzenegger and the smaller Danny DeVito. The scene starts with a close-up of DeVito, then moves to a long shot of DeVito and Schwarzenegger beginning a conversation at the edge of a parking lot. As the two converse, there are two 180-degree violations. The first is in a relatively straightforward cut between a long shot and a medium shot with the camera essentially perpendicular to the eyeline. In the shots immediately preceding the second 180-degree violation, the actors cross twice (see Figure 3), exchanging screen positions within close and medium shots. In shot 1 Schwarzenegger occupies screen-right, and looks to the left. When DeVito crosses in front of him, Schwarzenegger's gaze sweeps by the camera to establish a new gaze direction. Following a long shot, shot 2 is a medium shot showing Schwarzenegger saliently crossing behind DeVito, reestablishing their initial screen positions.

Once the two become stationary next to a car, the remainder of the conversation is shown in two medium/close shots. The first of these (shot 3, right before the violation shot) is a medium shot of Schwarzenegger towering over DeVito—not only is he clearly taller, but the camera is tilted up just a bit to further emphasize the height difference between the two. At this point, the camera dollies to the left towards the eyeline interrupted by a cut to a medium shot of DeVito that has crossed the eyeline, violating the 180-degree rule.

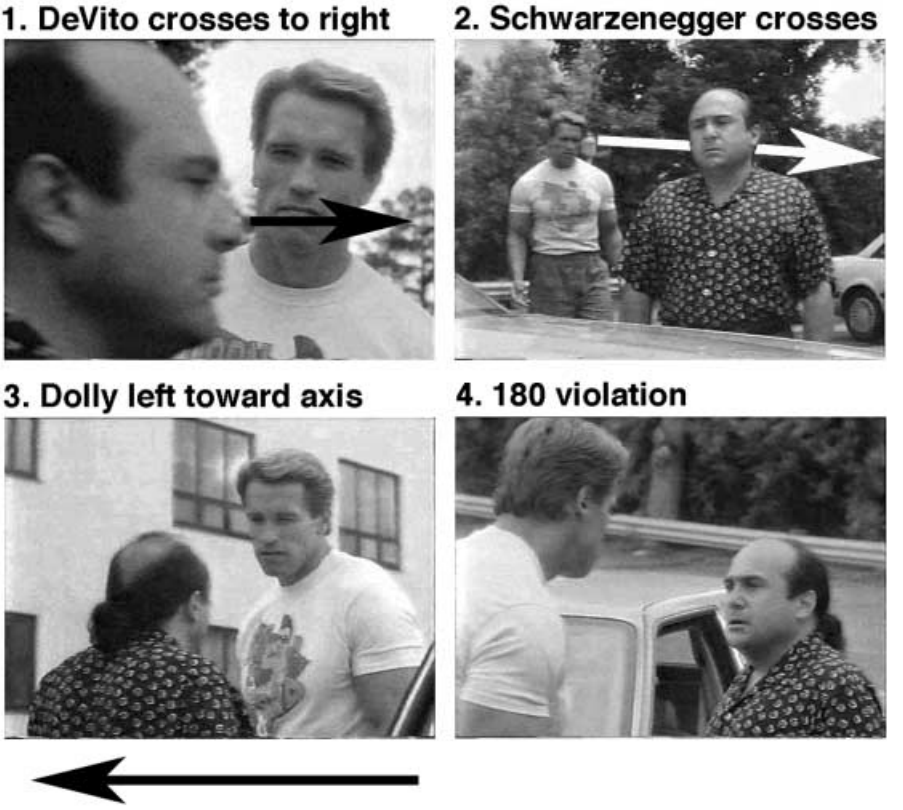


Figure 3. Shots preceding a 180-degree rule violation in the film Twins. *In shots 1 and 2, the actors change screen position twice. In shot 3, the camera begins to dolly to the left, approaching, but not reaching, the eyeline. In shot 4, the 180-degree rule is violated. On the cut, the camera is slowly dollying in.*

Several features of the post-violation shot appear to facilitate the cut. First, the camera is now raised slightly behind Schwarzenegger, looking down at DeVito. In addition, the leftward movement in the previous shot is continued by a subtle dolly into DeVito. The two camera movements create a broader movement around DeVito's face and move in to take a closer look at him. This movement reinforces the narrative as DeVito is now having a revelation in which the audience can feel involved.

This example is interesting for several reasons. Although the 180-degree rule is not inviolate, filmmakers do not appear to violate it casually. The two cross movements of the characters preceding the violation loosen the audience's expectations about the screen location, though both are rigorously consistent with the need to establish new eyelines with on-screen action. In this editing the filmmakers use space to reinforce the deeper meaning of the scene. The two crossing movements of the characters reveal how DeVito's self-absorbed emotional state leads him to look away and walk away from Schwarzenegger, who follows him, trying to counter his negative thoughts.

This scene could have employed an unbroken series of reverse angles, with the camera leading the two actors to the car, but this would forego a range of expressive spatial dynamics between the characters.

During the violation, several spatial cues appear to be emphasized as substitutes for the gaze direction as a source of information. The most basic of these are the two-shots that show both actors. Had the violating shot been a close-up, the audience would have been forced to rely solely on the now-unreliable cue of gaze direction. In addition, shots 3 and 4 combine to strongly establish the height differential between the characters, making the cut consistent with one relative spatial cue (gaze and height elevation), while it violates another (horizontal gaze direction). In addition, the two actors body positions are carefully replicated across the cut—Schwarzenegger leans forward in both shots, and DeVito stands back.

In addition to manipulating the balance between cues, the camera movements in the scene strongly facilitate a seamless cut. The initial dolly to the right in shot 3 happens after Schwarzenegger has suggested to DeVito that they can have a family and DeVito considers this possibility. The camera moves around DeVito as he rotates his head away from the audience, who now want to see his face more than ever. The 180-degree violation moves to a new camera position that represents a continuation of the initial dolly movement around DeVito to see his face, followed by a subtle movement in to get closer to DeVito to further emphasize his revelation. Thus, the 180-degree violation not only uses the specific spatial cues inherent to the medium shot from an external angle, the difference in gaze elevation, and the actor's body position, but it is reinforced by the narrative, which has induced the audience to ask questions that will be answered by the eyeline-violating view. (An external angle is a shot in which we see the back of one actor and the front of another while the two converse. This contrasts with an internal angle in which we see only one actor's face, while the other actor is entirely offscreen. See Arijon 1976.) One of the most interesting things about this example is how it demonstrates many of the principles of spatial representation that we discussed above. At the most basic level, this example demonstrates the centrality of gaze in spatial representation, but, more interesting, it relies on the dynamic reweighting of different spatial cues implied by the adaptive combination view. The key to this view is that there is a range of possible spatial cues, and that people can combine them in a way that is responsive to the current setting. In the case of the *Twins* example, the actors' gazes are clearly foundational cues, but they are not the only cues. Other cues such as the iconic significance of both actors, the height differential between them, and the flow of interest in the scene are also useful. It appears as though the director and editor have pushed us to rely less on gaze and more on the other cues by having the actors repeatedly exchange positions.

A Small-Scale Survey of the 180-Degree Rule

Of course, there are limits to the use of examples as they are not necessarily representative of the range of film editing practice, and they can easily be filtered to fit with preexisting hypotheses. In this section, we present a more systematic analysis of 180-degree rule usage, and also a preliminary experiment on the perception of space in films. For our survey, we follow Bordwell, Staiger, and Thompson (1985) who in their broad survey of classical (pre-1960) Hollywood cinema, observed that principles of spatial continuity (such as the 180-degree rule) were violated in only about 2 percent of cuts. Although Bordwell (2006) argues that there is substantial continuity between these earlier films and current practice, it is possible to argue that film editing practice has grown considerably more sophisticated and films are no longer so dependent on simple heuristics such as the 180-degree rule.

To further explore the 180-degree rule, we completed a small-scale systematic survey of editing practice in a current group of films that range in the level of art and craft expertise applied to them. We were particularly interested in whether there would be a substantial difference in the number of 180-degree violations, and the circumstances in which they occur, for well-known films of wide appeal as compared with more mediocre films. From one view, it is possible that minimally skilled filmmakers know so little that they have difficulty organizing camera positions during a shoot to conform with the 180-degree rule. In contrast, expert filmmakers may be able to more effectively organize scenes to avoid 180-degree violations. However, from our analysis of *Twins,* we can argue that there is a more interesting possibility: top filmmakers may have a much deeper understanding of the full range of cues that organize online spatial representations and can skillfully manipulate the salience of different spatial cues, lessening their reliance on simple heuristics such as the 180-degree rule.

In this section, we present a more systematic analysis of 180-degree rule usage, and also a preliminary experiment on the perception of space in films.

We were also interested in the circumstances under which violations occur. It is possible that experts' ability to effectively balance spatial cues leads their 180-degree violations to occur in different circumstances than for novices. As we have seen from our example, one very basic way of lessening the impact of 180-degree violations is to ensure that the violating view contains additional objects such as the targets of a character's gaze. Therefore, one might expect that experts are more likely to violate the 180-degree rule in medium shots, as in the *Twins* example, or, if they are going to violate on a close-up, to ensure that they are using an external angle.

As an initial test of these hypotheses we selected twenty films rated best and twenty films rated worst on the Internet Movie Database Top 250 and Bottom 100 lists. Of course, it would have been possible to choose the top

films from a list based on critical opinion, but we wanted to focus on films that were widely enjoyed and not necessarily appreciated for their groundbreaking distinctiveness. We chose three scenes from each film to analyze by randomly selecting time points within the film that were more than ten minutes prior to the end of the film. We chose scenes with at least two centers of interest and edits that might plausibly test the 180-degree rule.

In total, we analyzed sixty scenes from the best films and sixty scenes from the worst films. We found that 9 percent (11/120; 95% confidence interval ranges from 3.6% to 14.7%) of the scenes we analyzed contained 180-degree violations, and that there was no substantial difference in the proportion of violations in the best and worst films. The violation rate was 8 percent (5/60) in the bottom twenty, and 10 percent (6/60) in the top twenty. This clearly demonstrates that 180-degree violations are not extreme rarities. Conversely, those who claim that the 180-degree rule is an outdated oversimplification of editing practice would have to acknowledge that the vast majority of scenes in a wide range of films do conform to the rule. In fact, if one focuses on the proportion of shots that conform to the rule, it is clear that filmmakers violate the rule in only a very small proportion of cuts (this proportion is very similar to the violation rate observed by Bordwell, Staiger and Thompson [1985]).

We also looked at the specific circumstances under which violations occur to see if we could detect a substantial difference between the works of expert and novice filmmakers. To do this, we classified the shots that caused the violations by coverage (long, medium, or close), and noted whether they were external or internal shots. This examination revealed a very interesting contrast between the best and worst films. The violations in the best films tended to occur in medium shots, and, in the one case where the violation occurred in a close-up, it was an external angle. In contrast, three of the five violations in the worst films occurred in internal close-ups. This can be taken as evidence that the best filmmakers are at some level aware that the gaze-reliant 180-degree rule is but one of several interchangeable kinds of information that can be used to organize reverse-angle editing because they only violated the rule when other information was available to replace the now-inconsistent gaze cue.

Looking at stills of the violations in the best films reveals some additional principles that resonate strongly with research on spatial memory. The violations in both *Lord of the Rings: Return of the King* (2003) and *Once Upon a Time in the West* (1968) make conspicuous use of medium shots that show substantial elements of the immediate spatial geometry. In *Lord of the Rings,* the two characters stand at the mouth of a cave so that the shot preceding the violation has the cave in the background, and the shot after the violation has the outdoors in the background. In the violation from *Once Upon a Time,* not only is there a difference in gaze elevation between the characters, but the compo-

sition of the post-violation shot emphasizes the staircase supporting one of the characters. Thus, both of these cuts seem to rely on the fact that people readily encode the local geometry of spaces that encloses objects, people, and events and can use this as a substitute for an eyeline match. Other occurrences of this are easy to find. For example, in a scene from *Music and Lyrics,* the two main characters converse in a rectangular beauty shop. One faces into the back of the shop, and one faces out of it, so the view on the pre-violation shot depicts the back of the shop, while the view on the post-violation shot depicts the street outside the shop through the window. Thus, in deemphasizing gaze direction as a spatial cue, these shots seem to increase reliance on just the kind of geometric spatial cues that Hermer and Spelke (1994) found young children using.

Based on these preliminary observations, it is possible to hypothesize that expert filmmakers know something interesting and generalizable about how people perceive, encode, and create mental representations of visual space. Consistent with psychological research exploring representations of small-scale spaces, it appears that local geometry is a strong and useful cue in organizing views online. However, our observations of film editing practice suggest that the psychological research might be a bit too narrow in the local cues it explores. Most of this research has explored how the shape of a room and arrangement of objects and landmarks affect spatial representations. Recall how Hermer and Spelke observed children relying on the long axis of a rectangular room to find a hidden object. This and similar experiments (see, e.g., Epstein 2004; McNamara and Valiquette 2004) have explored how people seem to automatically encode room-shape, and must expend more effort encoding locations relative to landmarks and other nongeometric cues such as the wall color. In some views, these geometric cues are limited to a reliance on the specific visual form of a local environment. But if the intuitions embedded in expert 180-degree violations are relevant to these representations, we might need to expand the range of automatically encoded spatial cues to include familiar categorical distinctions such as inside/outside. This is reminiscent of language-inspired analyses of spatial perception (e.g., Colin 1995).

On this view, automatically encoded spatial axes can be organized around several different kinds of spatial information, including both local geometry and categorical directional contrasts. If we assume that these cues are used to create representational axes for local environments, then it becomes a relatively straightforward matter to identify one, or multiple useful axes in any given environment, and to explore how people select from the options available to them when encoding realistic spatial layouts. The fact that filmmakers seem to have discovered the range of possible view-organizing axes inherent to a broad range of environments suggests that research in spatial representation might not only explore more possible sources of spatial information,

but that it might also directly explore how people actually select from among multiple intrinsic axes available in any given scene.

The Role of Meaning in Representations of Space

One of the most interesting things that the *Twins* example makes clear is that the online coding of space is not just a question of representing a simple grid specifying the locations of objects using cues such as gaze and environmental geometry. This approach has some truth to it, but it misses one of the most interesting aspects of filmic space: it is controlled, dominated, and negotiated in a subtle dance of goals, emotions, and conflict. Murch (2001) argues that narrative consistency between shots is even more important than the spatial consistency implied by the 180-degree rule. However, even more interesting is the degree to which the narrative organizing a scene often converges with the locations of objects in the scene. This was suggested in our discussion of *Twins* where Danny DeVito's movement away from Arnold Schwarzenegger reinforces DeVito's self-focused thoughts, and Schwarzenegger's towering over DeVito while entreating him creates emotional tension. Examples of the meaningful use of space abound. For example, in the film *Local Hero* (1983), McIntyre, the protagonist, finds comfort and friendship when he leaves his home in the United States and travels to a small fishing village in Scotland. The contrast of the emotional distance between people in the United States and the closeness of the Scots is repeatedly emphasized using a whole series of spatial cues. In the United States conversations are often depicted using internal reverse angles showing one character at a time, in some cases separated by barriers such as glass walls or windows. In contrast, once McIntyre gets to Scotland, conversations occur in cramped spaces depicted using medium two-shots, and McIntyre is often forced into close no-barrier interactions with others.

The use of space to tell stories is frequently discussed in even the most basic film texts. It is, after all, one of the most powerful and distinctive tools filmmakers have at their disposal. In addition, a few basic empirical demonstrations have confirmed that specific interpersonal spatial relationships affect the emotional impact of a scene. For example, Kraft, Cantor, and Gottdiener (1991) systematically manipulated the relative elevations of two characters and observed that the higher one was, indeed, rated as more dominant. This gives some sense of the deep meaning of space, but it is still reduced to a matter of a relatively simple cue.

In a recent experiment we observed hints of similar interactions of space and meaning. We asked participants to view a series of short films in which one person showed an object to another person, and then placed it on a table. As illustrated in Figure 4, the sequence of shots started with an establishing shot showing the two actors from the side. (We will refer to the actor doing

Figure 4. Sequence of views in cut-on-gaze experiment. In the establishing shot the model begins a reach to an object. In the second shot, the model brings the object near her head, looks to the observer, and begins the hand movement to place the object on the table. In the third shot of the canonical sequence, the observer follows the object to the table with his eyes. The fourth shot begins when the object has just been placed on the table. In the reversed sequence, the object is placed on the table then the observer moves his eyes.

the showing—on the right in this particular film—as the model, and the actor doing the looking as the observer.) The second shot was a close-up of the model as she brought the object into view and then looked from the object to the observer. The model then began the motion of putting the object on the table. Just after the beginning of this motion, the view cut to the observer who was seen in an extreme close-up following the object to the table with his eyes. Once the observer's eyes had followed the object all the way to the table, the camera cut to a close-up of the object resting on the table with the model's hand withdrawing from it. The key to this sequence is that the only cue for the object's location on the table was the observer's gaze; all of the other shots were purposely ambiguous with regard to the location of the object. Not only did all of the sequences use the same shot of the model beginning to put the object on the table, but they also employed the exact same close-up of the object hitting the table. The only thing differentiating the apparent location of the object was where the observer looked. We tested whether participants could effectively use the observer's gaze to choose from among four possible locations on the table depicted in a still of the initial establishing shot. The locations depicted in the still were arranged in a square, with two to the left of the observer's midline and two to the right. Thus, participants needed to determine from the observer's close-up whether his gaze fell to a spot on the table that was to his left or right, and to determine

whether the spot was closer to the observer's side of the table or closer to the model's side. Because there were four possible locations for the object, we assessed the percentage of trials on which subjects correctly chose the location that the observer looked at and compared it with a 25 percent baseline of chance responding. As far as we know, only a few experiments have tested whether gaze alone allows participants an effective spatial representation of non-overlapping shots (Hochberg 1994). Research demonstrating the costs of testing from a viewpoint different from the learned viewpoint (Diwadkar and McNamara 1997) suggests that this task might not be trivial.

In total, participants viewed eight different films depicting these events with various actors. In half of the sequences, the shots were ordered in the canonical fashion described above. However, for half the order of the first close-up of the observer and the shot of the object on the table were reversed. The idea was to see whether accuracy would be lessened if participants saw a sequence that did not conform with the typical intentional interpersonal event in which A shows B an object and B immediately looks at it, an interaction typical of the kind of event that we use Theory of Mind to interpret. Overall, participants were able to determine the location of the object at above-change levels, but in our initial analysis, we were surprised to find that our ordering manipulation had very little effect on participants' accuracy in using gaze to identify the location of the objects.

Despite this disappointment, we were able to discover something very interesting based on a post-experiment questionnaire. We asked all the participants whether they had noticed any particular differences among the experimental films, and to specify what they were. Then, we supplied two sets of hand drawn stills depicting the canonical and reversed sequences, and asked participants whether they thought that all of the films followed one of the sequences, or whether there was a mix of sequences. We were surprised to find that most of the participants failed to realize that the films had different sequences of shots. To test whether this awareness of the differences between the sequences was associated with task performance, we focused on the eight out of forty-three participants who clearly recognized that both sequences had been in the experiment. We compared these eight with the other thirty-five participants. (It is important to note that a few of these participants claimed to recognize that there were different sequences, but failed to mention this in the open-ended questions, making it unclear whether they had actually noticed or not. To be conservative, we did not count them among the subjects who clearly had noticed.) As shown in Figure 5, there was no reliable difference between the aware and unaware participants for the canonical sequence, but for the reversed sequence, the aware participants were significantly more accurate in reporting the placement on the table ($t(41) = 2.568$, $p = .014$) than the nonaware participants.

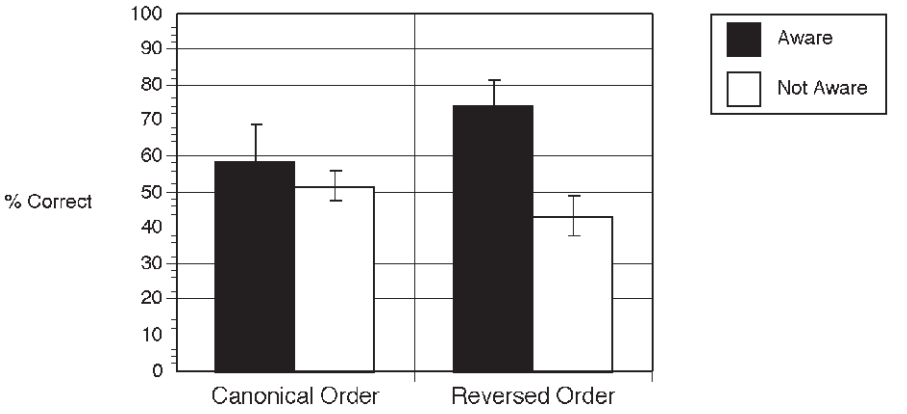


Figure 5. Results of cut-on-gaze experiment.

This effect was, at first, puzzling, but closer consideration of the meaning of the reversed films affords an interesting interpretation of this effect. Our initial idea was to create an "incorrect" stimulus that would be inconsistent with the cognitive framework participants would apply to encoding the space. However, when we reviewed the films it became clear that they were not so odd; there was a coherent, and interesting interpretation of the reversed film. In the normal film, the observers just appeared to follow the object to the table, but in the reversed film, the observers appeared to keep their focus on the model as she placed the object on the table. If one is thinking deeply about the reversed film, one gets the impression that the observer is suspicious of the model continuing to look at her while momentarily ignoring the distraction of the action the model is engaged in. In a broader sense a viewer might speculate that the observer is staying focused on the internal mental processes of the model, perhaps attempting to divine the model's motivations.

It appears as though some participants were not focused on the apparent meaning of the event, and therefore did reasonably well with the typical sequence, perhaps because it was easily coded using simple default sequence of inferences characteristic of the relatively automatic first-pass versions of Theory of Mind. When these defaults were violated, the non-noticing viewers did not focus closely enough to link the delayed downward gaze shift with the object, perhaps because making this link required them to expend the effort to hold a distinctive sequence in working memory. On the contrary, a minority of participants effectively tracked the meaning of the scene, which allowed them to accurately encode the location of the objects. This ability can be seen as evidence that something like the second nonautomatic stage of a dual process Theory of Mind was employed by this subset of participants.

A couple of caveats about our finding seem important. First, the default sequence has clearly not resulted in error-free performance. Accuracy on the spatial task ranged between 50 percent and 70 percent (relative to a 25% chance baseline); and, even when detecting whether the gaze landed to the

left or to the right of the observer, accuracy was still not at ceiling (70–80%). This inaccuracy may have occurred for several reasons; perhaps the most likely is that participants' coding was sometimes less precise than required by the test. Even for the apparently simple left-right contrast, a more general coding that the object is close to and in front of the observer would constitute a potentially information-rich representation that would not be helpful in succeeding on the location test, which asks for a more precise location. In any case, it would be important to determine the precision of default coding. One interesting possibility is that the default in a two-person conversation includes a clear left-right directional distinction (as suggested by the 180-degree rule), and that the default in coding the locations of people relative to objects is slightly more general, sometimes preserving only a rough estimate of proximity (or reachability) in front of the person. The second, perhaps more important, caveat is that most of the evidence for a putative default mode comes from tasks that require far more than the minimal "default" encoding. In our observer/model experiment, participants were repeatedly required to report the location of an object in a way that would not characterize normal film viewing. The only real evidence that the spatial coding involved can be considered a default comes from our argument (supported by some limited empirical evidence), and from the tradition in film craft that 180-degree violations are disruptive.

One reason why these qualifications are important (and actually of interest) is that filmic space is often not completely coherent or veridical. These departures from veridicality have been commented upon frequently; they include slight "cheats" in which actors or objects are repositioned between shots in ways that enhance dramatic effects at the expense of correct spatial information. Some of these spatial inconsistencies can be fairly dramatic. For example, many commentators have discussed how the film *Casablanca* (1942) depicts spaces such as Rick's club in views that are often contradictory, and sometimes downright strange when viewed closely (see, e.g., Kimersgaard 1998; see also Roger Ebert's commentary track on the *Casablanca* special edition DVD). If we are to argue that viewers code some spatial information as a matter of course, it would be helpful to use the spatial violations that are allowed to occur in film to develop a hypothesis about the limits to this coding. As implied above, one limit might be a question of the precision of viewers' representations. The cheat shots in *Casablanca* might be nondisruptive because viewers do code spatial information, but at a level of abstraction that encompasses both the real and cheated spaces. Another important possibility, implied by Murch (2001), is that people do, at some level, notice these inconsistencies but they are overwhelmed by consistency in narrative and emotional logic. There might even be momentary awareness of the conflict that is soon forgotten or discounted (Beck, Angelone, and Levin 2004). Al-

though it is difficult to test the aware-then-immediately-forgotten hypothesis, the idea of narrative consistency is more testable. For example, it would be possible to test for detection of spatial conflict, just as previous research has explored spontaneous (or the lack of) awareness of continuity errors (e.g., Levin and Simons 1997). Such tests could be done using films with strong and weak narrative coherence.

Much of this hypothesized elaborative support and use of meaning to encode space is new to cognitive theories of spatial perception, making these ideas examples of a potentially productive two-way interaction between cognitive science and film.

To summarize, in this section we have argued that spatial coding in film relies on the adaptive combination of spatial cues, and that basic (and perhaps automatic) use of spatial information is sometimes elaborated upon. As we observed in our experiment, this elaboration may be guided by Theory of Mind, as participants track distinctive mental contents (e.g., our observer's "suspicion" about the model) to coordinate the use of an atypical looking event as a spatial cue. To apply this to filmmaking practice, editors can probably rely on the audience to use a basic set of spatial cues, and even to flexibly weight them. However, if the audience is sufficiently involved, elaborative processes may amplify these basic encodings, to allow increased use of subtle and/or distinctive cues. It is important to note that much of this hypothesized elaborative support and use of meaning to encode space is new to cognitive theories of spatial perception, making these ideas examples of a potentially productive two-way interaction between cognitive science and film.

Long-Term Spatial Memory: The Transition from Short-Term Online Coding to Long-Term Memories

If 180-degree violations represent the online creation of mental representations of space, it is important to consider the fate of these representations: Do they disappear once a new scene is experienced, or are they retained and used to build a more complete long-term representation of larger multiple-location environments? On the one hand, research on narrative text comprehension suggests that the spatial representations inherent to multi-scene stories are only represented if the reader has a functional need to perceive them (Zwann and Radvansky 1998). Hochberg and Brooks (1996) have argued that spatial representations in film are only relevant locally (e.g., when understanding the relation of one shot to the next), and not necessarily integrated or remembered over a longer period. On the other hand, recent research on visual memory for objects in scenes does suggest that in the normal course of viewing a scene long-term memories build up (Hollingworth 2004). In addition, some researchers have suggested that visual narratives such as film may create stronger, more easily perceived spatial representations than text (Magliano, Miller, and Zwann 2001).

From a psychological point of view, representations of the layout of spaces in motion pictures are a useful test of hypotheses about long-term visual memory because they are part of the viewers' task to understand a visual narrative, while not suspecting that they will be tested with the location of any specific thing. In contrast, because the typical lab setting inevitably suggests that the participants will be tested, it is difficult to know whether they remember spatial information because they think that they must or simply from a natural process.

To explore long-term spatial representations, we sought settings where viewers have the maximum possible chance to create long-term mental representations of the layout of filmed scenes. We compared participants who had viewed many episodes of television shows that repeatedly depict single multiple-part sets with participants who had never seen the shows before (Levin, in review). In our first experiment, we borrowed a method from Shelton and McNamara (2001), and asked participants to imagine a target location on the set of the show *ER,* and to indicate in which direction one would have to point at the target from another location on the set. For example, as depicted in Figure 6, participants would first see a still of the target location, in this case the emergency room entrance, and then they would see a base location from which they used a set of arrows to point to the target. In each of our first two experiments we observed a striking failure of highly experienced viewers to remember the configuration of the *ER* set. Mean pointing error was quite high (only slightly better than chance responding), and there was no difference between these viewers and other participants who had never seen the show. It is important to note that this did not occur simply because the spatial pointing task was so difficult that it erased any possible difference between viewers and nonviewers. In our second experiment, we added a control condition in which all participants successfully used the same scale to imagine and point to locations that had just been shown to them outside the room where they were being tested.

Why do frequent viewers have so much difficulty remembering the layout of the *ER* set? One interesting possibility is that this particular set is difficult because it allows cameras to be placed anywhere, pointing in any direction. The show's producers went to considerable pains to create a full emergency room, and they rely heavily on Steadicam shots that swoop around the set, showing it from all points of view. This technique contrasts with other shows that use more theatrical sets and live audiences. In these cases, there is a "fourth wall" corresponding to the seating section in a traditional theater, and the set consists of one or more rooms defined by the three walls that the audience sees. In these sets all camera positions are near the unseen fourth wall, which prevents potential conflict between different spatial frames of reference.

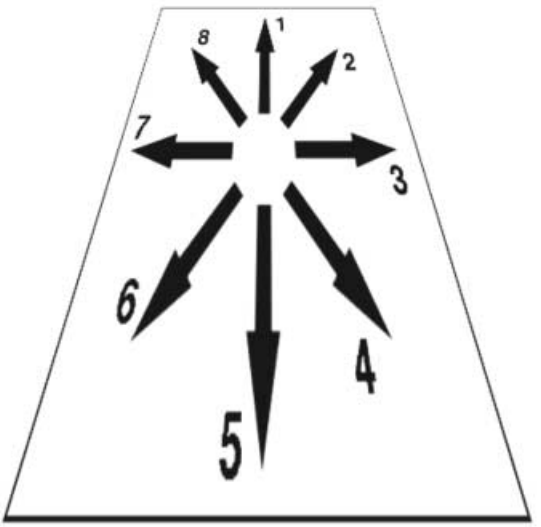
A

Which way is it TO
THE EMERGENCY ENTRANCE
from …

Figure 6. Sequence in layout memory experiment.

B

…from
THE STAIRS



Consider, for example, the *ER* still in the top of Figure 6 depicting the entrance to the emergency room. This location is where many dramatic sequences begin, as gravely injured patients are rushed into the hospital. One important aspect of this location is that the trauma rooms, where sequences end, are to screen-right. In fact, the vast majority of the set is to the right in this view. In some sense, there is considerable spatial consistency to these scenes—the paramedics always rush to the middle of the set. However, the camera is not constrained to cover the entrance from a position opposite it, near the admittance desk. Sometimes, the camera starts out near the door and follows behind the gurney as it enters the hospital. This situation might create a visual challenge for viewers because it invokes two conflicting reference frames (see McNamara and Valiquette 2004). One reference frame is an intrinsic environment-centered representation of the set itself. For the purposes of simplification, assume that this reference frame is organized around an axis running up and down the rectangular *ER* set perpendicular to the entrance door. In this reference frame the trauma rooms are always a left turn into the middle of the hospital. However, it is also possible for viewers to think about the turn in terms of an egocentric-reference frame, or even in terms of an intrinsic reference frame relative to the TV set. In the former case, the depicted turn is to the viewer's right, and in the case of the TV set, it is a turn to screen-right. Because the camera sometimes looks out of and sometimes looks into the main set, some entrances to the hospital involve movement to screen-right, but some movements in the same direction relative to the set are to screen-left/viewer-left. These conflicts do not occur in more theatrical sets where the camera position is more constrained.

To test whether the unconstrained set of *ER* was responsible for the poor performance of our frequent viewers, we ran another experiment testing multiple examples of unconstrained sets (*ER* and *West Wing*), and more constrained sets with traditional fourth walls (*Drew Carey, Friends,* and *Third Rock from the Sun*). In this experiment, a large group of subjects completed a more simple task with stills from all of these shows. Participants simply saw a single still with two arrows under it (one pointing left and one pointing right), and indicated which arrow pointed to a named off-screen location. Before completing the direction responses for the TV shows, participants again gave responses for locations within and surrounding their immediate environment, and, after, they indicated how frequently they watched each of the depicted shows. As shown in Figure 7, results demonstrated that viewership was considerably more effective for the constrained-view shows than for the multiple-view shows such as *ER* and *West Wing*.

These results are interesting because they hint at a potentially new hypothesis that could relate short-term spatial representations and more long-term representations. Based on our discussion of the 180-degree rule, it

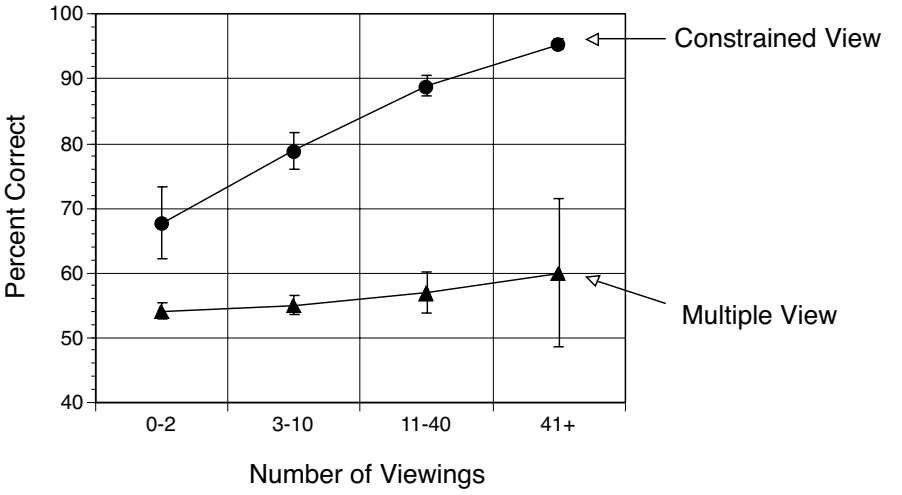


Figure 7. Results of layout memory experiment (Levin, in review). Percent correct for multiple-view and constrained-view shows.

appears likely that viewers often create online representations of the layout of important objects in a scene, and possibly the geometry of the environment containing them. The data described above suggests an interesting answer about their fate—sometimes they are not retained and integrated into a fuller model of an environment as a whole, but sometimes they are. When representational reference frames rely on consistent axes it may be considerably easier for viewers to relate more episodic representations of specific subsections of a set into a more complete representation as a whole. This idea might find support in previous research suggesting that changes in imagined perspective are difficult to track because they can create conflicts between allocentric (or object-to-object) representations (e.g., the diegetic space of the *ER* set) and egocentric (or person-to-object) representations (perhaps of the relation between the viewer and his/her television) (May 2004). One particularly interesting possibility is that this consistency is what allows a passively navigated setting such as a TV show to be learned. Research on spatial memory suggests that active navigation improves spatial representations because it allows people to update representations of a previous view to match with a current view from a different station point (see, e.g., Wang and Simons 1999). Perhaps one important part of this active updating process is that it affords navigators the possibility of overcoming conflicts between intrinsic environment-centered reference frames and other more egocentric-reference frames.

Of course, one reason why people might not integrate subspaces in the multiple-view sets is that these are not functionally relevant to understanding the story. This resonates with recent research on sentence comprehension suggesting that people do not automatically create spatial representations of texts, even when reading single sentences describing simple scenes. It is only

when the task makes clear to subjects that these relationships are functionally relevant that spatial information is encoded (Jahn 2004). However, it is important to note that if space is nonfunctional in multiple-view sets, there is no reason to think that this makes them different from the consistent-view sets, which do seem to benefit from integration of subspaces.

In this section, we have argued that there is something interesting about the process of combining short-term spatial representations of mediated subspaces into long term representations of complex environments. In particular, it appears that subspace representations can be combined if the representational axes organizing the spaces are consistent with each other. In this case, the cognitive research lays the foundations for understanding audience representations by establishing general ideas about spatial reference frames, but our analysis of film can elaborate how these reference frames can affect the combination of episodic representations in viewers who are engaging in a naturalistic spatial task that does not focus their attention solely on spatial encoding.

Conclusions

In this article we have explored short- and long-term spatial representations, and argued that there are close links between research in cognitive psychology and film editing practice. We agree with authors who argue that film studies and cognitive science are natural partners (Anderson 1996; Carroll 1996; Messaris 1994; Prince 1994), and have attempted to develop specific hypotheses that arise from this interaction. Our first hypothesis derived from editing practice surrounding the 180-degree rule, and basic research on spatial representation. Our analysis of editing practice and cognitive research first converged to suggest that there are a number of spatial cues including gaze, geometry, static properties that vary in the degree to which they are generally coded by default. However, both the cognitive research and editing practice converge to suggest that these defaults can be modified, both by long-term experience and by short-term priorities and cue validities. A key insight from editing practice is that the range of available cues may be wider than previously appreciated in the cognitive research.

Our second hypothesis attempts to describe ways in which meaning interacts with space. Following dual-process explanations of Theory of Mind, we suggest that some spatial processes relate things at specific locations in the environment to people's internal thoughts in a relatively automatic way, but other more complex thought-object relationships require deliberation. These more complex relationships are not necessarily coded by all viewers.

Our final hypothesis reflects the longer-term coding and storage of spatial information. It appears as though online coding sometimes results in longer-term representations, but sometimes does not. The variable distinguishing

these situations is the degree to which viewpoints are consistent—representations of sets where any view was possible appeared quite weak and resistant to experience and memory.

The goal of this paper was to review commonalities between film studies and cognitive psychology. We hope that these hypotheses provide not only a specific form to this interaction, but also a basis for new research on the cognitive processes underlying perception in both film and the real world. Even at the start of film scholarship, Münsterberg noted that film editing reflected a kind of internal mental play in which "the turning of our attention" to the objects and spaces on the screen was parallel to our attention to the "life around us" (2001:177). There is a fundamental way in which recent research has linked the turning of attention in film to that in the real world, both when considering the representation of visual properties and visual space. In both of these domains it seems likely that theories about film might benefit from grounding in cognitive and perceptual processes. However, we suspect that the impact of film theory on cognitive science might be equally interesting, as it can help us bring meaning, emotion, and even drama to what has previously been a relatively dry analysis of visual attention and representation.

Daniel T. Levin is currently an Associate Professor of Psychology in the Department of Psychology and Human Development at Vanderbilt University. His research explores the interface between cognition and perception. His work has been supported by grants from the NSF and NIMH. Dr. Levin won a Divison III APA New Investigator Award (2001), and his research has been frequently cited, described in introductory texts in cognition, perception, and general psychology, and covered in the mass media.

Caryn Wang is currently an undergraduate at Vanderbilt University majoring in cognitive science with an interest in film perception.

Acknowledgment

The authors thank Mark May and John Bateman for reading and commenting on this manuscript.

References

Anderson, Joseph, D. 1996. *The Reality of Illusion: An Ecological Approach to Cognitive Film Theory.* Carbondale: Southern Illinois University Press.

Apperly, I.A., K.J. Riggs, A. Simpson, C. Chiavarino, C., D. and D. Samson. 2006. "Is Belief Reasoning Automatic?" *Psychological Science* 17: 841–46.

Arijon, Daniel. 1976. *Grammar of the Film Language.* Los Angeles: Silman-James Press.

Baron-Cohen, Simon. 1995. *Mindblindness: An Essay on Autism and Theory of Mind.* Cambridge, MA: Bradford Books/MIT Press.

Beck, Melissa R., Bonnie L. Angelone, and Daniel T. Levin. 2004. "Knowledge about the Probability of Change Affects Change Detection Performance." *Journal of Experimental Psychology: Human Perception and Performance* 4: 778–91.

Bordwell, David. 2006. *The Way Hollywood Tells It*. Los Angeles: University of California Press.

Bordwell, David, Janet Staiger, and Kristin Thompson. 1985. *The Classical Hollywood Cinema: Film Style and Mode of Production to 1960.* New York: Columbia University Press.

Butler, Samantha C., Albert J. Caron, and Rechele Brooks. 2000. "Infant Understanding of the Referential Nature of Looking." *Journal of Cognition and Development* 1: 359–77.

Butterworth, George, and Jarrett, Nancy. 1991. "What Minds Have in Common Is Space: Spatial Mechanisms Serving Joint Visual Attention in Infancy." *British Journal of Developmental Psychology* 9: 55–72.

Carroll, Noël. 1996. "Prospects for Film Theory: A Personal Assessment." Pp 37–70 in *Post-Theory: Reconstructing Film Studies,* eds. David Bordwell and Noël Carroll. Madison: University of Wisconsin Press.

Colin, Michael. 1995. "Film Semiology as a Cognitive Science. " Pp 87–112 in *The Film Spectator: From Sign to Mind,* ed. Warren Buckland. Amsterdam, Netherlands: Amsterdam University Press.

Diwadkar, Vaibhav A., and Timothy P. McNamara. 1997. "Viewpoint Dependence in Scene Recognition." *Psychological Science* 8 (4): 302–7.

Epstein, Russell A. 2005. "The Cortical Basis of Scene Processing." *Visual Cognition* 12: 954–78.

Flavell, John H., Frances L. Green, and Elanor R. Flavell. 1990. "Developmental Changes in Young Children's Knowledge about the Mind." *Cognitive Development* 5 (1): 1–27.

Frith, Uta and Joselyn E. Robinson. 1975. "Perceiving the Language of Films." *Perception* 4: 97–103.

Gopnik, Alison, Virginia Slaughter, and Andres Meltzoff. 1994. "Changing Your Views: How Understanding Visual Perception Can Lead to a New Theory of the Mind." Pp. 157–81 in *Origins of an Understanding of Mind,* eds. Charlie Lewis and Peter Mitchell. Hillsdale, NJ: Erlbaum.

Hermer, Linda, and Elizabeth S. Spelke. 1994. "A Geometric Process for Spatial Reorientation in Young Children." *Nature* 370: 57–59.

———. 1996. "Modularity and Development: The Case of Spatial Reorientation." *Cognition* 61: 195–232.

Hermer-Vazquez, Linda, Elizabeth S. Spelke, and A.S. Katsnelson. 1999. "Sources of Flexibility in Human Cognition: Dual-Task Studies of Space and Language." *Cognitive Psychology* 39: 3–36.

Hochberg, Julian. 1986. "Representation of Motion and Space in Video and Cinematic Displays." Pp. 22.1–22.84 in *Handbook of Perception and Human Performance, Volume 1: Sensory Processes and Perception,* eds. Kenneth R. Boff, Lloyd Kaufman, and James P. Thomas. New York: John Wiley and Sons.

———. 1994. "Binding 'What' and 'Where' across Successive Views: Theories, Models, Tests."

Paper presented at the annual meeting of the Psychonomic Society. St Louis, MO.
Hochberg, Julian, and Virginia Brooks. 1996. "Movies in the Mind's Eye." Pp 368–87 in *Post-Theory: Reconstructing Film Studies,* eds. David Bordwell and Noël Carroll. Madison: University of Wisconsin Press.
Hollingworth, Andrew. 2004. "Constructing Visual Representations of Natural Scenes: The Roles of Short- And Long-Term Visual Memory." *Journal of Experimental Psychology: Human Perception and Performance* 30: 519–37.
Intraub, Helene, and Michael Richardson. 1989. "Wide-Angle Memories of Close-up Scenes." *Journal of Experimental Psychology: Learning, Memory and Cognition* 15: 179–87.
Jahn, Georg. 2004. "Three Turtles in Danger: Spontaneous Construction of Causally Relevant Spatial Situation Models." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30: 969–87.
Kimersgaard, Lars B. 1998. "Editing in the Depth of the Surface: Some Principles of Graphic Editing." *P.O.V.* 6: 75–90.
Kraft, Robert N. 1987. "Rules and Strategies of Visual Narratives." *Perceptual and Motor Skills* 64: 3–14.
Kraft, Robert N., P. Cantor, and C. Gottdiener. 1991. "The Coherence of Visual Narratives." *Communication Research* 18: 601–16.
Lee, Sang A., Anna Shusterman, and Elizabeth S. Spelk. 2006. "Reorientation and Landmark-guided Search by Young Children: Evidence for Two Systems." Psychological Science, 17: 577–582.
Leslie, Alan M., Ori Friedman, and Tim P. German. 2004. "Core Mechanisms in 'Theory of Mind." *Trends in Cognitive Sciences* 8: 528–33.
Levin, Daniel T. In review. "Spatial Representations of the Sets of Familiar and Unfamiliar Television Programs."
Levin, Daniel T., and Daniel J. Simons. 1997. "Failure to Detect Changes to Attended Objects in Motion Pictures." *Psychonomic Bulletin and Review* 4: 501–6.
———. 2000. "Fragmentation and Continuity in Motion Pictures and the Real World." *Media Psychology* 2: 357–80.
Magliano, Joseph P., Jason Miller, and Rolf A. Zwann. 2001. "Indexing Space and Time in Film Understanding." *Applied Cognitive Psychology* 15: 533–45.
Mandler, Jean M., Dale Seegmiller, and Jeanne Day. 1977. "On the Coding of Spatial Information." *Memory and Cognition* 5: 10–16.
May, Mark. 2004. "Imaginal Perspective Switches in Remembered Environments: Transformation versus Interference Accounts." *Cognitive Psychology* 48: 163–204.
McNamara, Timothy, and Christine Valiquette. 2004. "Remembering Where Things Are." Pp. 3–24 in *Human Spatial Memory: Remembering Where,* ed. Gary L. Allen. London, UK: Lawrence Erlbaum.
Messaris, Paul. 1994. *Visual Literacy: Images, Mind, and Reality.* Boulder, CO: Westview Press.
Moll, Heinricke, and Michael Tomasello. 2004. "12- and 18-Month-Old Infants Follow Gaze to Spaces Behind Barriers." *Developmental Science* 7: F1–F9.
Münsterberg, Hugo. [1915] 2001. "Why We Go to the Movies." Pp. 171–82 in *Hugo Münsterberg in Film: The Photoplay—a Psychological Study and Other Writings,* ed. Alan Langdale. New York: Taylor and Francis Group.
Murch, Walter. 2001. *In the Blink of an Eye: A Perspective on Film Editing.* Los Angeles: Silman-James Press.
Newcombe, Nora S., and J. Sluzenski. 2004. "Starting Points and Change in Early Spatial Development." Pp. 25–40 in *Human Spatial Memory: Remembering Where,* ed. Gary L. Allen. London, UK: Lawrence Erlbaum.

O'Regan, Kevin. 1992. "Solving the 'Real' Mysteries of Visual Perception: The World as an Outside Memory." *Canadian Journal of Psychology* 46: 461–88.

Prince, Stephen. 1994. "The Discourse of Pictures: Iconicity and Film Studies." *Film Quarterly* 47: 16–28.

Ratliff, Kristin R., and Nora S. Newcombe. 2008. "Is Language Necessary for Human Spatial Orientation? Reconsidering Evidence from Dual Task Paradigms." *Cognitive Psychology* 56: 142–63.

Reisz, Karl, and Gavin Miller. 1953. *The Technique of Film Editing*. New York: Hastings House.

Rensink, Ronald A. 2002. "Change Detection." *Annual Review of Psychology* 53: 245–77.

Schulman, Arthur I. 1973. "Recognition Memory and the Recall of Spatial Location." *Memory and Cognition* 1: 256–60.

Sedgwick, Hal A. 1982. "Visual Modes of Spatial Orientation." Pp. 3–34 in *Spatial Abilities: Development and Physiological Foundations,* ed. Michael Potegal. New York: Academic Press.

Shelton, Amy L., and Timothy P. McNamara. 2001. "Systems of Spatial Reference in Human Memory." *Cognitive Psychology* 43: 274–310.

Shoqueirat, Mohammed A., and Andrew R. Mayes. 1991. "Disproportionate Incidental Spatial-Memory and Recall Deficits in Amnesia." *Neuropsychologia* 29: 749–69.

Simons, Daniel J. 1996. "In Sight, out of Mind: When Object Representations Fail." *Psychological Science* 7: 301–5.

———. 2000. "Current Approaches to Change Blindness." *Visual Cognition* 7: 1–15.

Simons, Daniel J., and Daniel T. Levin. 1997. "Failure to Detect Changes to People During Real-World Interactions." *Psychonomic Bulletin and Review* 5: 644–49.

Wang, Frances, and Daniel J. Simons. 1999. "Active and Passive Scene Recognition across Views." *Cognition* 70: 191–210.

Woodward, Amanda L. 2005. "Infants' Understanding of the Actions Involved in Joint Attention." Pp. 229–62 in *Advances in Child Development and Behavior, Volume 33,* ed. Robert Kail. Oxford: Elsevier.

Zwann, Rolf A., and Gabriel A. Radvansky. 1998. "Situation Models in Language Comprehension and Memory." *Psychological Bulletin* 123: 162–85.

Filmology

Curtiz, Michael. 1942. *Casablanca.* USA.

Forsyth, Bill. 1983. *Local Hero.* USA.

Jackson, Peter. 2003. *Lord of the Rings: Return of the King.* USA.

Lawrence, Mark. 2007. *Music and Lyrics.* USA.

Leone, Sergio. 1968. *Once Upon a Time in the West.* Italy.

Reitman, Ivan. 1988. *Twins.* USA.

Van Sandt, Gus. 1989. *Drugstore Cowboy.* USA.