Title
Changing Viewpoints During Dynamic Events

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 27(27)

ISSN
1069-7977

Authors
Garsoffky, Barbel
Huff, Markus
Schwan, Stephan

Publication Date
2005

Peer reviewed

Changing Viewpoints During Dynamic Events

Bärbel Garsoffky (b.garsoffky@iwm-kmrc.de)
Knowledge Media Research Center, Konrad-Adenauer-Str. 40
72072 Tübingen, Germany

Markus Huff (m.huff@iwm-kmrc.de)
Knowledge Media Research Center, Konrad-Adenauer-Str. 40
72072 Tübingen, Germany

Stephan Schwan (s.schwan@iwm-kmrc.de)
Knowledge Media Research Center, Konrad-Adenauer-Str. 40
72072 Tübingen, Germany

Abstract

After observing dynamic events memory performance seems to be viewpoint dependent. The main idea of the experiment was if special viewing conditions could weaken this viewpoint dependency. To test that in a learning phase short clips of dynamic basketball scenes were presented. This presentation showed the whole basketball scene from one viewpoint, or the viewpoint changed during the presentation, at which the two viewpoints were connected either by a moving camera or by a cut. In the test phase visual recognition from familiar and unfamiliar viewpoints was tested. Results showed that (a) the presentation modes in the learning phase differed in the way that recognition was best after presenting only one viewpoint, and it was worst if two viewpoints were connected by a cut. (b) recognition was viewpoint dependent, and that this viewpoint dependency did not disappear when the observer was forced already in the learning phase to understand a viewpoint change.

Keywords: Psychology; memory; representation.

Viewpoints on Dynamic Events

During visual recognition of dynamic events, a viewpoint deviation effect can be observed. That is observers will recognize an event better, if they see the initial presentation of the event and the presentation of the test stimulus from the same viewpoint. Recognition will get worse, if the viewpoints between learning phase and memory test phase differ; this could be shown for dynamic events like soccer episodes or dynamic ball scenes (Garsoffky, Schwan, & Hesse, 2002, 2004) and was also found for the visual recognition of static objects (e.g. Tarr, 1995), and static scenes (Diwadkar & McNamara, 1997).

The present study deals with the question whether this viewpoint deviation effect diminishes, if the presentation of a dynamic event makes use of specific presentation strategies, thereby leading to a viewpoint-independent and therefore more flexible cognitive representation. Previous research e.g. showed, that the use of canonical viewpoints at least weakens the viewpoint deviation effect during recognition (Garsoffky et al., 2004). Also as could be shown by Garsoffky et al. (2002, experiment 2) for dynamic events and by Diwadkar and McNamara (1997) for static scenes, the use of more than one viewpoint during the initial presentation leads to a multiple viewpoint representation. Nevertheless in these cases the representation does not become viewpoint independent in the sense that the representation generalizes to novel viewpoints and points in time (Garsoffky et al., 2002, experiment 3).

Additionally there are hints that if several viewpoints on the same scene are realized the way of viewpoint change may play an important role: It seems that some kind of cognitive spatial updating during self movement of the observer of a static scene can lead to a viewpoint independent (Simons & Wang, 1998) or at least to an orientation independent representation (Sholl & Nolin, 1997). Sun, Chan and Campos (2004) also found that memory performance was less viewpoint dependent if the learning process was active instead of passive. In general, visual recognition is better if the observer can actively choose viewpoints in the learning phase (James, Humphrey, & Vilis, 2002; Wraga, Creem-Regehr, & Proffitt, 2004). Similarly Christou and Bülthoff (1999) found a more viewpoint independent representation, if movement was possible during the initial presentation of a static scene in a virtual environment – independent if the movement was actively controlled by the observer or managed by the program. Furthermore Christou, Tjan and Bülthoff (2003) observed that extrinsic cues (in their experiment a realistic room as background) on how the viewpoint changed between an initial learning phase and a later test phase helped during shape recognition. These results indicate that the presentation mode of a viewpoint change can become important for the development of a viewpoint independent representation at least for static scenes.

But how does an observer deal with varying viewpoints? How does s/he transfer from one viewpoint to another one? There are hints from two areas: On the one hand multiple viewpoints research on static objects forces participants to think of the same object from various viewpoints and suggests alignment processes that can either be more "discrete" or more "analogue"; discrete means that the effort of aligning one viewpoint with another one is independent of the distance in degrees between the two viewpoints, whereas analogue stands for processes that become the more effortful the more the two viewpoints differ (e.g. Shepard & Metzler, 1971; Ullman, 1989). On the other hand designer of visual media as e.g. movie makers often deal with the problem how to combine various viewpoints on the same scene (Arijon, 1991; Bordwell & Thompson, 1993). They developed rules and realized various ways to solve the problem. In the case of a movie e.g. one camera viewpoint can be transferred into another or connected to another camera viewpoint by a “filmic cut”, i.e. one camera position immediately follows another camera position, or by a "pan", i.e. a static camera that continuously turns, or by a "move", i.e. the camera moves around the scene and thereby also continuously shows the scene from the various viewpoints. Empirical literature is ambiguous whether all of these filmic strategies are well adapted to the cognitive processes of observers: Some studies showed that e.g. filmic cuts are understood even by completely media inexperienced people (Messaris, 1994), whereas other researchers postulate that only continuously changing viewpoints reflect the biological possible everyday experience of humans and therefore should be understood better than abrupt viewpoint changes (Gibson, 1982). Supporting this latter idea Kipper (1986) found that recall and recognition of a static scene was better if during the initial presentation of the scene the transition between various viewpoints was realized by a moving camera instead of simply adding segments with different viewpoints directly together.

On the one hand several findings from static scenes could also be replicated for dynamic scenes, indicating some commonalities in spatial representation of static and dynamic scenes (Garsoffky et al., 2002). But on the other hand dealing now with questions of presentation mode, the observer´s situation strongly differs between the reception of a static or a dynamic scene: In contrast to static scenes the observer has to deal with two concurrent sources of dynamic change – namely the dynamic inherent in the scene (e.g. movement of actors or objects) and the dynamic of the presentation mode (e.g. a moving camera). This overlap of dynamics may pose additional computational load on the observer.

Looking at the cognitive task to transfer one viewpoint into another one the following experiment examines if filmic advice as cuts or moving cameras can influence necessary cognitive processes (Salomon, 1994) when aligning two different viewpoints.

Experiment

The following study presented dynamic events, namely dynamic basketball scenes. The players of the two teams could be discriminated by the colour and pattern of their tricots, and it could always be seen how one team made a basket (see figure 1). In the experiment, in a learning phase participants always saw a dynamic basketball scene and then in a test phase had to recognize video stills from this event, i.e. visual recognition was tested. To look for the viewpoint deviation effect (Garsoffky et al., 2002) the viewpoints between the initial presentation of the dynamic basketball scene and the test phase either differed or not. It was expected, that recognition would be best, if the viewpoint did not change from the learning to the test phase, and that it would become worse, if the viewpoint of the video still in the test phase differed from the viewpoint in the clip presented in the initial learning phase.

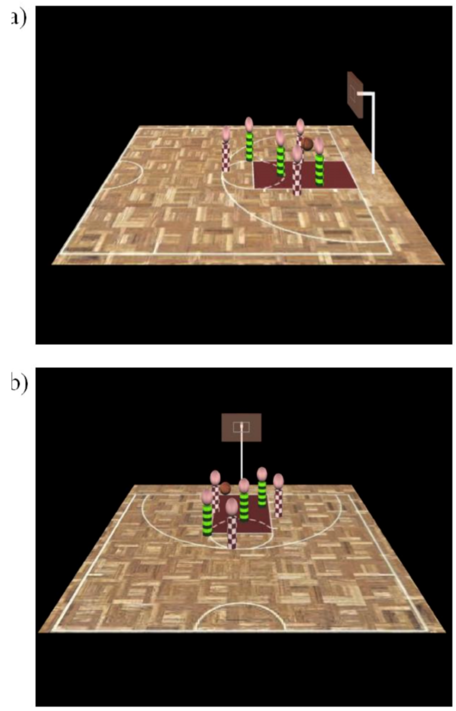

Figure 1: Section of a basketball scene seen from a) sideline and b) middleline camera.

The main goal of the experiment was to test, if special forms of presenting the dynamic basketball scene in the learning phase could weaken the viewpoint deviation effect. The idea was to change the viewpoint already in the initial learning phase so that the observer would be forced to transfer one viewpoint into another, and additionally to realize the connection of the two different viewpoints in the learning phase by the use of different filmic means, namely cuts or moving cameras. Looking at everyday life where observers are often moving (e.g. walking, travelling in cars), and taking into account the amount of time we spend consuming visual media as Television or movies (using lots of moving cameras and cuts) it was expected that if a viewpoint change had to be understood already during the initial presentation, the observer would build a cognitive representation of the basketball scene that would be more viewpoint independent (i.e. the viewpoint deviation effect should become weaker) than if the learning phase only used one viewpoint. Furthermore this viewpoint independency should be stronger if the connection between the two viewpoints in the learning phase supported cognitive processes of the observer when transferring one viewpoint into another one; i.e. the filmic advice should be variably qualified in weakening the viewpoint deviation effect. At least it was also expected that in general recognition would be worse if viewpoint changes during the in initial presentation occured, because of the double cognitive load emerging from the dynamic inherent in the scene (moving players) *and* the variation of viewpoint (cut or move).

Method

Participants Four male and eight female students from the university of Tuebingen participated in this experiment. They were paid for their participation.

Apparatus Experimental procedures were controlled by an IBM computer and realized by MediaLab and directRT. The basketball clips were presented on a black background in the middle of a 18″ monitor with a resolution of 800 x 600 px.

Stimulus material and design Twenty-four dynamic basketball scenes were programmed using 3ds max. Each scene showed players from two teams, which could be differentiated by the colors and patterns of the tricots; the duration of each scene was 10 sec. Each scene ended with one team making a basket, and it was attempted to design the dynamic basketball scenes as realistic as possible. For each dynamic basketball scene three presentation modes were realized: The whole scene was presented either from one viewpoint or from two viewpoints, i.e. the viewpoint changed during the presentation of the scene. In the case of two viewpoints the connection between them was either abrupt by a filmic cut or continuous by a moving camera. The moving camera started after 4 sec of total scene duration and lasted 2 sec (see figure 2). Also presentation was controlled for viewpoint position (sideline vs. middleline camera), sequences of viewpoint position (first sideline then middleline camera or vice versa), and for the main direction of players movement (left vs. right). In addition some training scenes were programmed, following the same variations as described for the experimental scenes.

To measure recognition, for each scene video stills were made, that either presented a cutout of the dynamic basketball scene seen in the initial learning phase or distractors, i.e. video stills presenting the same players but another move. The video stills stemmed from three points in time, namely after 20%, 50% and 80% of total scene duration (i.e. after 2, 5, and 8 sec), and presented the scene from varying viewpoints. Please refer to figure 3 for all camera positions as exemplified for a basketball scene with players moving to the right basket which is presented from sideline camera in the learning phase: 0° is the viewpoint of sideline camera (in this example the learning viewpoint) and is in the test phase the viewpoint of video stills with no deviation from the learning viewpoint; 90° is middleline viewpoint and the viewpoint of video stills with 90° deviation from learning viewpoint; 45° and 135° are the viewpoints for video stills with 45° and 135° degree deviation from the learning viewpoint.

These variations resulted in a design with the variables "presentation mode" during the initial presentation (no change of viewpoint, two viewpoints connected by a moving camera, two viewpoints connected by an abrupt cut), "viewpoint deviation" (0°, 45°, 90° or 135° deviation between the viewpoint seen in the initial scene presentation at that point of time in the scene and viewpoint of the video still item used in the recognition test phase), and "point of time" (video still shows a point of time after 2 sec scene duration, i.e. 2 sec before onset of the filmic mean, after 5 sec scene duration, i.e. 1 sec after onset of the filmic mean, or after 8 sec scene duration, i.e. at least 2 sec after end of filmic mean).

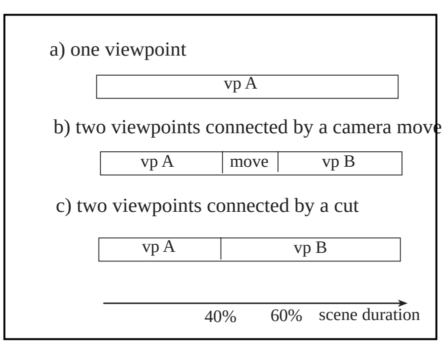

Figure 2: The three presentation modes during the initial presentation: a) one viewpoint (vp) throughout the whole scene (i.e. sideline or middleline camera), b) changing viewpoint (i.e. from sideline to middleline camera or vice versa) and connecting the two viewpoints by a camera move, and c) changing viewpoint and connecting the two by a cut.

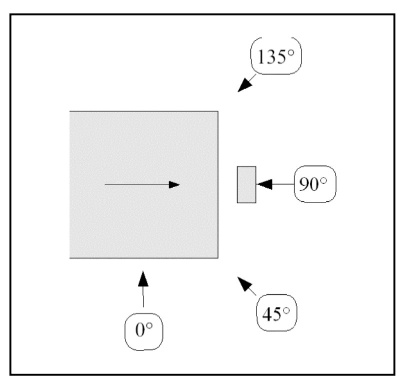

Figure 3: Camera viewpoints e.g. for a move to the right basket seen from sideline camera in the learning phase.

Procedure All participants were tested individually and received written instructions to the main part of the experiment – namely a description of the stimulus material and their recognition task. First they passed through a training phase, the data of which were not analyzed. The experimental phase encompassed 24 dynamic basketball scenes, i.e. 24 blocks. Each block consisted of an initial learning phase followed by a test phase. In the learning phase participants saw a basketball scene from either one single viewpoint (sideline or middleline camera), with changing viewpoint during the presentation (from sideline to middlleline camera or vice versa) realized by either a moving camera or a cut. One second later, they successively saw 24 video stills: twelve video stills presenting the originally seen dynamic basketball scene (three points of time of the scene each presented from four different viewpoints) as well as twelve distractor video stills which used the same viewpoints but presented other basketball scenes, i.e. the scenes showed the same players (same colors) but the video stills stemmed from other moves. So to perform the recognition task participants had to decide, if a video still showed a moment of the move seen before in the film or another move. The order of the video stills was randomized. Each video still stayed on the screen until the participant pressed one of two buttons (one marked with "j" for the german word "ja" which means "yes", and one marked with "n" for the german word "nein" which means "no"). After the participant had reacted to a video still there always was a short delay of one second before the next video still was presented. The order of blocks (i.e. the different basketball scenes and the different conditions according to one or two viewpoints) was randomized and each scene was presented in the learning phase to a third of the participants from only one viewpoint, to another third of the participants with changing viewpoint connected by a camera move, and to the other third of participants with changing viewpoint connected by a cut. Every participant saw each basketball scene only one time, i.e. under only one condition (one viewpoint / two viewpoints connected by a move / two viewpoints connected by a cut). But across all participants every dynamic basketball scene was presented under each condition.

Results

Recognition accuracy For each participant his or her number of "hits" (the number of video stills correctly recognized as showing a moment from the basketball scene which he or she had previously seen) was determined. Across all participants and conditions a mean of 65 % hits resulted. An ANOVA with repeated measurement was performed, including the variables "presentation mode" (one viewpoint, change of viewpoint by a move or by a cut; within subjects), "viewpoint deviation" (0˚, 45˚, 90 or 135˚; within subjects), and "point of time" (after 2, 5, or 8 sec scene duration; within subjects). A significant main effect for "presentation mode" was found ($F(2,22) = 5{,}03$, MSE = 0,05145, $p = .016$) with 69.5 % hits when there was only one viewpoint during the initial scene presentation, 64.8 % hits when there were two viewpoints connected with a moving camera, and 61.1 % hits when there were two viewpoints connected by a cut (see figure 4); single comparisons according to Scheffé revealed a significant difference (5%) between the viewing condition with one viewpoint and the condition with two viewpoints connected by a cut. Furthermore the variable "viewpoint deviation" was significant ($F(3, 33) = 5{,}26$, MSE = 0,01556, $p = .004$) with 69.2 % hits at 0˚ viewpoint deviation between learning and test phase, 63.7 % hits at 45˚ viewpoint deviation, 63.1 % hits at 90˚ viewpoint deviation, and 64.6 % hits at 135˚ viewpoint deviation. There was a significant linear trend for this variable ($p = .018$). Also the variable "point of time" was significant ($F(2,22) = 10{,}42$, MSE = 0,07792, $p = .001$) with 57.6 % hits after 2 sec scene duration, 65.3 % hits after 5 sec scene duration, and 72.6 % hits after 8 sec scene duration. This variable too was significant linear ($p = .001$). There were no significant interactions.

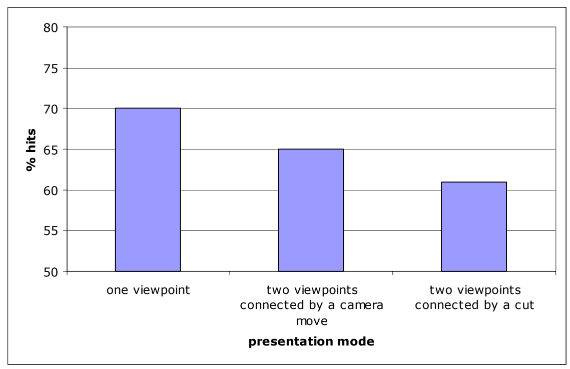

Figure 4: Recognition accuracy depends on presentation mode.

Speed of recognition As a second dependent variable, reaction time was measured, i.e. the lapse of time from the beginning of each video still presentation until the participant pressed either the "j"- or the "n"-button. There was no significant correlation between % hits and reaction times ($r = -0.144$), i.e. there is no speed-accuracy-trade off, and both measures can be interpreted. The following analysis only accounted for reaction times (RTs) to "hits" (i.e. correct "j"-reactions). Extreme RTs above 10 sec were excluded. This resulted in an exclusion of 0,8% of all RTs. To exclude outliers from analysis is a common method when dealing with reaction times (e.g. Cameron & Frieske, 1994; Diwadkar & McNamara, 1997; Eley, 1982) because extremely slow responses indicate lapses of a participant´s attention on a particular trial. Then an ANOVA with repeated measurement was performed, including the variables "viewpoint deviation" (0˚, 45˚, 90 or 135˚; within subjects), "presentation mode" (one viewpoint, two viewpoints connected by a move or two viewpoints connected by a cut; within subjects), and "point of time" (after 2, 5, or 8 sec scene duration; within subjects). A significant main effect for "viewpoint deviation" was found (F (3, 30) = 5,01, MSE = 385663,462, p = .006) with 2055 ms at 0˚ viewpoint deviation between learning and test phase, 2030 ms at 45˚ viewpoint deviation, 2163 ms at 90˚ viewpoint deviation, and 2337 ms at 135˚ viewpoint deviation (see figure 5). There was a significant linear trend for this variable (p = .045). Again the variable "point of time" was significant (F (2,20) = 5,25, MSE = 1287665,5, p = .015) with 2334 ms after 2 sec scene duration, 2210 ms after 5 sec scene duration, and 1895 ms after 8 sec scene duration. This variable as well was significant linear (p = .039). There were no more significant main effects or interactions.

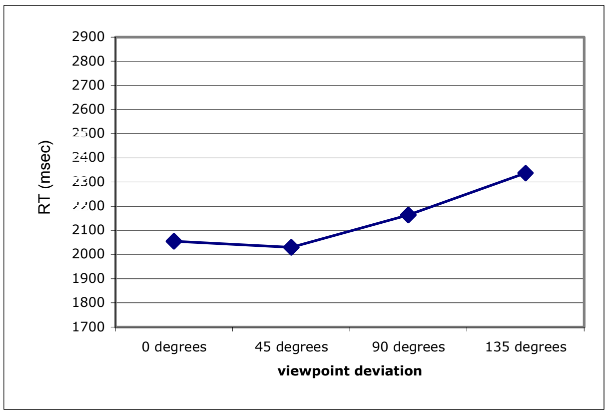

Figure 5: Speed of recognition depends on viewpoint deviation.

Discussion

In the present study once again (Garsoffky et al., 2002, 2004) the viewpoint deviation effect occurred. Accuracy and speed of recognition of dynamic events depend on the viewpoint deviation between initial learning phase and later test phase: Recognition became worse when the viewpoint of the video still used as test stimulus differed from the viewpoint during the initial presentation of the basketball scene. This again shows that the cognitive representation of a dynamic event seems to be viewpoint dependent and that later memory retrieval processes have to align new and familiar viewpoints.

The main idea of the experiment was to test if the visualization of changing viewpoints reduces this viewpoint dependency. To test this question two presentation modes were used during the initial basketball scene presentation: One viewpoint was connected to another viewpoint either by an abrupt cut or by a moving camera. These different modes of presentation indeed differed: (a) alltogether recognition accuracy was best, when there was no camera change or camera movement during the initial scene presentation; i.e. a viewpoint change in addition to the dynamic inherent in the scene leads to a suboptimal cognitive representation, what might be attributed on higher cognitive load during scene perception. (b) when the viewpoint changed in the film clip, recognition was worse if the camera viewpoint changed abrupt than when the camera moved to present the scene from another viewpoint. This can be seen as a hint, that when the viewpoint changes, a moving camera makes it easier to build a coherent cognitive representation, so that later memory retrieval is better than when the viewpoint changes abruptly – a result that is in line with our everyday life experience, where we mostly change viewpoint in a smooth fashion (Gibson, 1982; Kipper, 1986).

But although the two utilized presentation modes seem to differ in their cognitive demands, they both do not reduce the viewpoint deviation effect during recognition, i.e. there was no significant interaction between “presentation mode” and “viewpoint deviation”, neither for accuracy nor for speed of recognition. When the viewpoint changed during the initial presentation of the scene, the cognitive representation stored every part of the scene dependent from the viewpoint that was used in the film clip during this part of the event – irrespective how the change between viewpoints was realized. That is even if a viewpoint change within a scene presentation is realized in a way that is easier to understand (moving camera instead of cut) the cognitive representation does not become more flexible, in contrast to the findings of Christou and Bülthoff (1999), viewers do not benefit from continuous motion, maybe due to the overlapping of the sources of change (moving players and moving camera). For dynamic scenes, as could be shown in our experiment, every single section of an event is still stored viewpoint dependent, although the viewpoint can change in the cognitive representation of an event e.g. between the first half and the second half of the event (what again is in line with findings in Garsoffky et al., 2002).

Additionally as in previous experiments with soccer episodes (Garsoffky et al., 2002) again a recency effect was found for speed and accuracy of recognition of dynamic events: Later moments of the dynamic basketball scene were recognized faster and better than earlier moments. This was independent from the presentation mode, i.e. even if the viewpoint changed during the presentation of the scene, this recency effect was stable.

So again the viewpoint deviation effect for dynamic events proves to be very robust: Even if participants are forced to understand a viewpoint change during the initial presentation of a scene, and even if this change is formed in a familiar way, the cognitive representation remains viewpoint dependent, and participants depend on the original seen viewpoint during memory retrieval.

Acknowledgments

The work on this experiment was supported by grant from the DFG (Deutsche Forschungsgemeinschaft). The authors thank Wolfgang Schmidt for his help programming the stimulus material.

References

Arijon, D. (1991). *Grammar of the film language*. Los Angeles: Silman-James Press.

Bordwell, D. & Thompson, K. (1993). *Film art: An introduction* (4th ed.). New York: McGraw-Hill, Inc.

Cameron, G. T. & Frieske, D. A. (1994). The time needed to answer: Measurement of memory response latency. In A. Land (ed.), *Measuring psychological responses to media*. Hillsdale, NJ: Lawrence Erlbaum, pp. 149 – 164.

Christou, C. G. & Bülthoff, H. H. (1999). View dependence in scene recognition after active learning. *Memory & Cognition*, 27 (6), 996 – 1007.

Christou, C. G., Tjan B. S., & Bülthoff, H. H. (2003). Extrinsic cues aid shape recognition from novel viewpoints. *Journal of Vision*, 3, 183 – 198.

Diwadkar, V. A. & McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science*, 8 (4), 302 - 307.

Eley, M. G. (1982). Identifying rotated letter-like symbols. *Memory & Cognition*, 10 (1), 25 - 32.

Garsoffky, B., Schwan, S., & Hesse, F. W. (2002). Viewpoint dependency in the recognition of dynamic scenes. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 28 (6), 1035 – 1050.

Garsoffky, B., Schwan, S., & Hesse, F. W. (2004). Does the viewpoint deviation effect diminish if canonical viewpoints are used for the presentation of dynamic sequences? *Proceedings of the 26th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 428 - 433.

Gibson, J. J. (1982). *Wahrnehmung und Umwelt: Der ökologische Ansatz in der visuellen Wahrnehmung* (transl. ed.). München, Wien, Baltimore: Urban und Schwarzenberg.

James, K. H., Humphrey, G. K., & Vilis, T. (2002). "Active" and "passive" learning of three-dimensional object structure within an immersive virtual reality environment. *Behavior Research Methods, instruments & Computers*, 34 (3), 383 – 390.

Kipper, P. (1986). Television camera movement as a source of perceptual information. *Journal of Broadcasting & Electronic Media*, 30 (3), 295 – 307.

Messaris, P. (1994). *Visual literacy: Image, mind, and reality*. Boulder, San Francisco, Oxford: Westview Press.

Salomon, G. (1994). *Interaction of media, cognition, and learning* (2nd ed.). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Shepard, R. N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701 - 703.

Sholl, M. J. & Nolin, T. L. (1997). Orientation specificity in representations of place. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23 (6), 1494 - 1507.

Simons, D. J. & Wang, R. F. (1998). Perceiving real-world viewpoint changes. *Psychological Science*, 9 (4), 315 - 320.

Sun, H.-J., Chan, G. S. W., & Campos, J. L. (2004). Active navigation and orientation-free spatial representations. *Memory & Cognition*, 32 (1), 51 – 71.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review*, 2 (1), 55 - 82.

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, 32, 193 - 254.

Wraga, M., Creem-Regehr, S. H., & Proffitt, D. R. (2004). Spatial updating of virtual displays during self- and display rotation. *Memory & Cognition*, 32 (3), 399 – 415.