COGNITION

Cognition 70 (1999) 191–210

Active and passive scene recognition across views

Ranxiao Frances Wang[a], Daniel J. Simons[b,*]

[a]*Massachusetts Institute of Technology, Cambridge, MA, USA*
[b]*Department of Psychology, Harvard University, 33 Kirkland Street, Room 820,*
Cambridge, MA 02138, USA

Received 20 May 1998; accepted 9 February 1999

Abstract

Recent evidence suggests that scene recognition across views is impaired when an array of objects rotates relative to a stationary observer, but not when the observer moves relative to a stationary display [Simons, D.J., Wang, R.F., 1998. Perceiving real-world viewpoint changes. Psychological Science 9, 315–320]. The experiments in this report examine whether the relatively poorer performance by stationary observers across view changes results from a lack of perceptual information for the rotation or from the lack of active control of the perspective change, both of which are present for viewpoint changes. Three experiments compared performance when observers passively experienced the view change and when they actively caused the change. Even with visual information and active control over the display rotation, change detection performance was still worse for orientation changes than for viewpoint changes. These findings suggest that observers can update a viewer-centered representation of a scene when they move to a different viewing position, but such updating does not occur during display rotations even with visual and motor information for the magnitude of the change. This experimental approach, using arrays of real objects rather than computer displays of isolated individual objects, can shed light on mechanisms that allow accurate recognition despite changes in the observer's position and orientation. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Scene recognition; Rotation; Viewer-centered representation

1. Introduction

Real-world object and scene recognition faces a fundamental problem: the

---

* Corresponding author.

PII: S0010-0277(99)00012-8

retinal projection of the environment changes whenever the observer or objects in the environment move. Changes to the relative positions of the observer and objects can lead to size and orientation changes in the retinal projection of the environment. Yet our visual system somehow finds stability in these changing images. Two distinct approaches to achieving stability across view changes have been proposed in the literature. The system may selectively encode features of the scene that are invariant to perspective changes and use those features in object and scene recognition. For example, we may represent the object-centered spatial relationships among the parts of an object. Alternatively, our system may employ some transformation rules to compensate for changes in the retinal projection, thereby providing a common basis for comparing two different views. For example, we may mentally rotate an object until it is aligned with a previous representation or we could interpolate between two or more views to recognize objects from new perspectives.

Research on object recognition across views has provided some support for each of these possibilities. For example, Biederman and colleagues (Ellis et al., 1989; Biederman and Cooper, 1991, 1992; Cooper et al., 1992; see also Bartram, 1976) used a priming paradigm and measured the response latency to name line drawings of familiar objects. In their studies, the amount of priming was unaffected by changes in the retinal size of the object from study to test (scaling invariance). Furthermore, naming latency was impervious to changes to the position of the object in the visual field and to the object's orientation in depth. Biederman and Gerhardstein (1993) showed similar orientation invariance when observers were asked to match individual shapes (Geons), name familiar objects, and classify unfamiliar objects.

In contrast, many other studies suggest that object recognition performance is view-dependent; recognition accuracy and latency differ as the test views deviate from the studied view (e.g. Shepard and Metzler, 1971; Shepard and Cooper, 1982; Rock et al., 1989). With wire-frame or blob-like objects in same-different judgment tasks (Bülthoff and Edelman, 1992; Tarr, 1995; Tarr et al., 1997), subjects typically show fast, accurate recognition for test views within a small distance of the studied view and impaired performance for novel views. Furthermore, the impairment seems to be systematically related to the magnitude of the difference between studied and tested views, particularly for changes to the in-depth orientation of an object. The greater the rotation in depth away from the studied view, the longer the response latency (see also Tarr and Pinker, 1989). Such findings necessarily imply that object representations are viewer-centered.

Another critical piece of evidence in support of viewer-centered representations is that when two or more views of the same object are provided at study, subjects subsequently generalize to intermediate views but not to other views (Bülthoff and Edelman, 1992; Kourtzi and Shiffrar, 1997). A number of models of object recognition have attempted to account for this finding by positing mechanisms that operate on viewer-centered representations. For example, linear combinations of 2D views (Ullman and Basri, 1991) and view approximation (Poggio and Edelman, 1990; Vetter et al., 1995) are both consistent with these data. However, in order to interpolate between two or more views, the initial views must first be linked to the same object. That is, subjects must recognize that the same object is being viewed in the first and second studied views even though those views differ. It is unclear from these models how this initial matching is accomplished, particularly if the views are relatively far apart and the object is not symmetrical (see Vetter and Poggio, 1994). Although these models may not fully account for the nature of object recognition for novel views, the empirical data seems to support the claim that representations of individual objects are view-dependent.

Both view-independent and view-dependent models of object recognition seem to capture some aspects of how the visual system accommodates view changes. For example, when the learning period is relatively long and the object is relatively complicated and difficult to name, recognition may rely on viewer-centered representations. On the other hand, when objects are made of distinct parts whose spatial relationship can be coded easily, and when the task concerns more abstract knowledge such as naming or classification, recognition may rely on view-independent representations. Nevertheless, studies comparing these models typically test recognition for isolated objects and they ignore extra-retinal information that is available in real-world object recognition. Thus, neither model is likely to explain all aspects of object representation.

2. Recognition of object arrays

Recently, several laboratories have begun to consider the recognition of more complex, naturalistic displays (e.g. spatial layouts of objects) across views. Spatial layout representations are important for a number of reasons. First, most real-world object recognition occurs in the context of other objects rather than in isolation. Thus, it seems reasonable to study spatial layout representations to gain a clearer picture of the sorts of representations we might need from one view to the next. Second, developmental evidence suggests that spatiotemporal representations are robust early in infancy and that they may be the core of infant object representations until late in the first year of life (Spelke et al., 1995; Xu and Carey, 1996). Third, the ability to detect changes to the configuration or layout of a set of objects appears to be more robust to verbal interference than the ability to detect changes to individual object features (Simons, 1996a). Animals also seem to be sensitive to changes to the spatial layout of objects in the environment (e.g. Thinus-Blanc et al., 1992). Taken together, these findings raise the possibility that layout representations may be central to our experience of a continuous, stable visual world. We seem to use such information early in life and we can retain it across views more effectively than other sorts of information.

Despite the importance of spatial layout recognition to our understanding of object and scene representation, relatively few studies have directly examined the effect of view changes on layout perception. Several recent studies examined recognition of spatial layouts across views and found performance resembling that for single objects (Diwadkar and McNamara, 1997; see also Simons, 1996b; Shelton and McNamara, 1997; Wang and Simons, 1997; Simons and Wang, 1998). For example, response latency increases linearly as the angular distance between the studied view and tested view increases (Nakatani et al., 1996; Diwadkar and McNamara, 1997). Furthermore, when more views are presented during study, response latencies are predicted by the angular distance between the test view and the nearest studied view (Diwadkar and McNamara, 1997).

3. A hint from spatial reasoning studies

Although studies of spatial layout recognition are closer to real-world recognition, most have neglected an important source of information that may be central to real-world object and scene recognition. In real environments, observers have available many sources of information in addition to the retinal projection of the scene. For example, they have visual, vestibular, and proprioceptive information for their own movements. Such extra-retinal information may specify the magnitude of a change in view, thereby allowing view-independent performance and possibly view-independent representations.

Extra-retinal information usually differs for observer movements and object rotations, and studies of spatial representation and spatial reasoning find differences in performance for these types of change (e.g. Huttenlocher and Presson, 1973, 1979; Presson, 1982; Rieser et al., 1994; Amorim and Stucchi, 1997; Farrell and Robertson, 1998; Wraga et al., submitted). When observers are asked to specify where an object would be relative to a different viewing position, performance is affected by whether observers actually move or whether they imagine the position or orientation change. For example, when viewers are blindfolded and rotated to a new orientation, they point with little error to the objects previously seen even though they had never experienced the test perspective. However, when participants imagine themselves rotated by same amount, they are much slower and less accurate in their ability to point to where the object would be from the new perspective (Rieser et al., 1994; Farrell and Robertson, 1998; Wang, 1998). These results suggest that the representation of spatial layout around an observer is orientation-dependent. Furthermore, transformations of the representation depend on the nature of the information available about the perspective changes. The representation can be more easily transformed, or updated, when observers experience the sensory input corresponding to a perspective change, but not when the transformation is imagined.

Even for imagined perspective changes, performance is affected by the type of imagined transformation. For example, when children imagine themselves moving around a table-top display, they have great difficulty pointing to where one of the toys would be relative to their imagined position. In contrast, their performance is much better when they imagine the table rotating by the same amount (Huttenlocher and Presson, 1973, 1979). Adults show a similar difference in the ability to imagine self-rotation and display rotation (e.g. Presson, 1982). Strikingly, when the task requires predicting object positions, performance is better for imagined observer movement; for example, children are better able to judge which object would be to their left when they imagine themselves moving than when they imagine the array rotating (Huttenlocher and Presson, 1973, 1979). The two types of rotations still differ, but the relative ease of the tasks is reversed. This pattern of results suggests that pointing tasks and item judgment tasks may require different sorts of transformations of spatial representations: some require egocentric updating and others do not. More interestingly for present purposes, these differences in performance for imagined display and observer rotations raise the possibility that these two types of transformation may operate in the recognition of scenes across actual shifts in view and not just in imaginary ones. Recent evidence supports this possibility: recognition of real-world displays is easier after physical movements of the viewer than after actual display rotations of the same magnitude (Simons and Wang, 1998).

4. Scene recognition in real world

Despite evidence that imagined observer and display rotations lead to differences in performance, only recently has work in object and scene recognition considered this difference. Studies of object recognition have relied exclusively on display rotations to study view changes. This neglect of observer movement can be traced to the assumption that equivalent retinal projection changes should produce equivalent mental transformations of the visual representation. Because the retinal projection caused by a display rotation can be equivalent to that caused by an observer movement, display orientation changes have often been referred to as viewpoint changes even though the observer does not change viewing position. However, a recent series of studies showed that observer viewpoint changes have different effects on recognition of object arrays than do display orientation changes, suggesting that extra-retinal information for the position of the viewer is incorporated into the scene recognition process (Simons and Wang, 1998).

These studies (Simons and Wang, 1998) tested subjects’ accuracy in detecting a change in an array of objects in the real world following viewpoint and orientation changes. When observers moved to a different viewing position, they were able to detect changes to the spatial layout readily and accurately despite a view change of 50 degrees. However, when the observer remained stationary and the spatial layout rotated by 50 degrees, recognition performance was significantly disrupted. This difference in performance seemed to be tied to the nature of the transformation process and not to other differences in the nature of the view change. One such possible factor is the relation between the observer and other parts of the testing space. When observers move to a different viewing position, their view of other aspects of the environment changes (e.g. markings on the wall), not just their view of the particular display. In contrast, when a display is rotated relative to a stationary observer, other aspects of the scene remain constant. As a result, the spatial relationship between the display and its background changes. However, the effect of the surrounding environment cannot account for the difference in performance for viewpoint and orientation changes because when observers were tested in a dark room with phosphorescent objects the pattern of performance was essentially the same. Thus, accurate performance for observer movements does not rely extensively on the existence of environmental cues as a reference frame. In contrast, when information for self-motion was disrupted, performance was significantly impaired, suggesting that the representation subserving superior performance for viewpoint changes is not environment-centered. The representation must be viewer-centered and the difference between observer and display movements results from a difference in the nature of the transformation. Apparently, view-dependent layout representations are transformed or updated using extra-retinal information to account for observer movements.

5. Mechanisms of updating

Studies of navigation have shown that extra-retinal information can be used in updating one’s own position. Spatial representations of position and orientation rely on vestibular signals (e.g. Israel et al., 1996), proprioceptive and kinesthetic cues (e.g. Loomis et al., 1993; Berthoz et al., 1995), optical flow (Ronacher and Wehner, 1995; Srinivasan et al., 1996), magnetic fields (Frier et al., 1996), and energy expenditure (Kirchner and Braun, 1994). By using one or more of these sources of information, we can also compute the position and orientation of an object as we move or turn (e.g. Wehner and Srinivasan, 1981; Amorim et al., 1997). Despite the abundant information for the use of extra-retinal information in navigation and the differences in performance for imaginary display and observer movements, little is known about the updating mechanisms underlying the representation of spatial layout.

In principle, transformation mechanisms used to adjust our spatial representations to accommodate view changes could take one of two forms. One possibility is that a common view-transformation system underlies all view changes. Such a system would compute the expected retinal image using extra-retinal information about the amount of perspective change, regardless of whether the change is caused by observer movement or array rotation. For example, visual and proprioceptive input could be used to facilitate rotation of a mental image. Alternatively, a specialized system may represent viewer-to-object relationships and only update these representations using information about observer movements. Such updating would not occur when the observer is stationary.

Our previous studies seem to support the latter mechanism because they failed to show updating during orientation changes. However, the two types of view changes used in those experiments were different in two fundamental respects. Although observer viewpoint changes provided direct visual information about the magnitude of the view change, display orientation changes did not. Perhaps this direct visual information for the change in view facilitates updating or mental rotation. Another, potentially more important difference between the two types of view change is that observers actively caused the viewpoint changes but not the orientation changes. Performance is often more accurate under active control conditions. For example, adaptation when wearing distorting prisms seems to require active control of motion (e.g. Held and Hein, 1958; Held and Bossom, 1961; Held and Freedman, 1963). In fact, a number of studies have directly examined representations of spatial information under active and passive viewing conditions. In one study, infants were shown a toy being hidden. They were then either carried to a position on the opposite side of the room or they were allowed to move themselves to the new position. In a subsequent search for the toy, infants who actively moved were more successful (Benson and Uzgiris, 1985; although see also Acredolo et al., 1984). Children also provide more accurate distance estimations following active exploration of a space (Poag et al., 1983). Furthermore, a number of studies comparing active movement through a space with passive viewing of the space have found superior spatial knowledge for subjects who control their own movements (Peruch et al., 1995). Active observers are better able to estimate future positions (Larish and Andersen, 1995), to generalize to novel views of the same scenes (Christou and Bülthoff, 1997), and to navigate through the space (Gale et al., 1990).

Superior performance for viewpoint changes may result from active control over the view change; observers can update their representations for changes in their viewing position because their active control over the change provides additional information. Subjects viewing orientation changes lack active control over the view change. In a sense, this active/passive distinction is related to the notion of efference-copy (e.g. Kelso, 1977; Bridgeman, 1986; Bridgeman and Graziano, 1989; Bridgeman and Stark, 1991). When performing an action, a copy of the motor plan may be incorporated into the representation, allowing the observer to adjust representations for the change as it occurs. The experiments in this report examine whether active control of the view change accounts for the difference between viewpoint and orientation changes. By eliminating this important difference between the two types of view change, these studies directly examine differences in the transformation process. If performance for display rotations and observer movements is equal when observers have active control in both (or neither) conditions, then differences between the conditions cannot be attributed to distinct transformation processes. Instead, they would appear to rely on a common view transformation mechanism. However, if performance still differs when observers have active control over the change in both (or neither) conditions, the study would provide substantially stronger evidence for the operation of different transformation mechanisms for the two types of view change.

The first experiment replicates our previous studies in a completely within-subject design with one important refinement: observers were given visual information for the magnitude of the rotation in the orientation change condition. This manipulation guaranteed that subjects in both view change conditions would have visual information about the occurrence and magnitude of view changes, thereby eliminating one important difference between orientation and viewpoint changes. The second experiment compares performance for orientation changes when observers passively experience the view change and when they actively control the change. The third experiment examines the role of active and passive control of movement for viewpoint changes.

6. Experiment 1

This experiment served as a replication of earlier work comparing orientation and viewpoint changes (Simons and Wang, 1998), and tested the possibility that the availability of additional visual information would allow updating during orientation changes. Observers viewed layouts of real objects on a rotating table and were asked to detect changes to the position of one of the objects. We examined performance on this task across both shifts in the observer viewing position and rotations of the display. In all cases, visual information for the magnitude of the view change was available to observers.

6.1. Method

6.1.1. Participants

Sixteen undergraduate students participated in the study. Each received $7 as compensation.

6.1.2. Apparatus

The experimental display consisted of five ordinary objects (brush, mug, tape dispenser, box, rock) placed on five of nine possible positions on a rotating circular table (1.22 m diameter, 58 cm high). The positions on the table were arranged so that no more than two objects would be aligned with the observer's view from any of the viewing angles used in the experiments. A 1.8 m high screen occluded the table and the array of objects from the observer's standing position. Two observation windows (6.35 cm wide by 8.9 cm high) were positioned approximately 60 cm apart (1.04 m above the ground) and were covered by opaque curtains that could be lifted. The viewing windows were each approximately 90 cm from the center of the table. A pole affixed to the table extended through a narrow horizontal strip in the occluder and was visible from the observer's side of the display (see Fig. 1).

6.1.3. Procedure

On each trial, observers viewed a layout of the five objects on the table for 3 s through one of the viewing windows (Study Period). They then lowered the curtain and waited for 7 s. During the delay interval, the experimenter moved one of the five objects to a previously unoccupied position. Subjects then viewed the array again and indicated on a response sheet which object they thought had moved (Test Period).

Each subject experienced four different kinds of trials, 20 trials of each for a total of 80 trials. For half of the trials, observers remained at the same viewing window for both the Study and Test period (Same Viewing Position). For 20 of those trials, the experimenter rotated the table by 40 degrees during the delay interval (Different View). Observers could view the rotation as it happened by watching the rod that extended through the slot in the occluding screen. For the other 20 of those trials, the table was not rotated, so the observer's view of the display was the same (Same View).

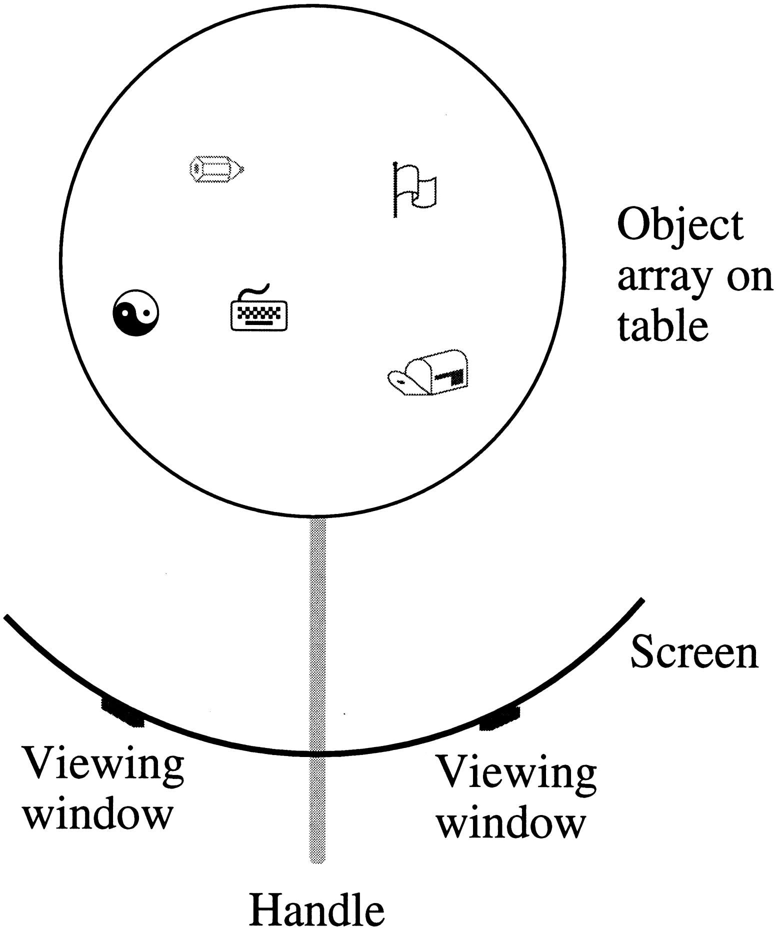

Fig. 1. An overhead view of the display.

On the other 40 trials, observers started the trial at one viewing window. During the delay interval, they walked to the other observation point (Different Viewing Position). For 20 of these trials, the experimenter did not rotate the table (Different View). As a result, observers experienced a 40 degree view change comparable to the view change in the Same Viewing Position condition. On the other 20 trials, the table rotated in the same direction and by the same magnitude as the observer movement so that the view of the table was the same at Study and Test (Same view). In summary, the four conditions included an orientation change (observer stationary, table rotates), a viewpoint change (observer moves, table stationary), and two identical view conditions (observer stationary, table stationary and observer moves, table rotates). For all trials in all conditions, observers were told prior to the Study period whether or not the table would rotate.

A preliminary study in which all of the trials were intermixed in a random order for each subject indicated that observers had difficulty switching among the task requirements for all four conditions. That is, they became confused as to whether the table would rotate or whether they should move. This confusion may have interfered with their strategies for performing the task. Although presenting each condition in a separate block of 20 trials would eliminate such confusion, order effects might influence the results. In order to avoid such order effects, trials were partially blocked and the order of blocks was counterbalanced both within and across subjects. More specifically, trials were arranged into blocks of five trials from the same condition, and these blocks of five were arranged into blocks of 20 trials with five trials from each condition. Four different orders of the conditions within a block of 20 were created using a Latin-Squares design, and each subject experienced all four of these blocks of 20. The order of the blocks of 20 was counterbalanced across subjects, also using a Latin-Squares design.

6.2. Results

This experiment can be thought of as a 2 (observer moves/observer stationary) × 2 (table rotates/table stationary) within-subjects design. As in the previous studies (Simons and Wang, 1998), performance was more disrupted by view changes caused by display rotations than view changes caused by observer movements (see Fig. 2). In fact, we found a reliable interaction between the observer viewing position (stationary or moving) and the view of the layout (same or different), $F(1,15) = 54.950$, $P < 0.0001$. When subjects remained at the same observation point throughout the trial, they were more accurate when they received the same view (i.e. the table was stationary) than when they received a different view (i.e. the table rotated), $t(15) = 6.296$, $P < 0.0001$. In contrast, when observers changed observation positions during a trial, they were more accurate when they received a different view (i.e. the table was stationary) than when they received the same view (i.e. the table rotated), $t(15) = 4.020$, $P = 0.0011$. Furthermore, observers were significantly more accurate when view changes were caused by their own movement (viewpoint change) than when they were caused by the table rotating (orientation change), $t(15) = 4.357$, $P = 0.0006$. When observers received a different view because they moved to a different viewing position, they were slightly less accurate than when they remained in the same viewing position throughout a trial and the table did not move, $t(15) = 3.048$, $P = 0.0081$. Although performance was impaired following either sort of view change, shifts caused by observer movements were far less disruptive than those caused by display rotations. Strikingly, when observers moved and the table rotated to provide the same view at Study and Test, performance was no better than when the array rotated in front of a stationary observer, $t(15) = 0.339$, $P = 0.9876$.

6.3. Discussion

The results of this experiment replicated the central finding from previous work

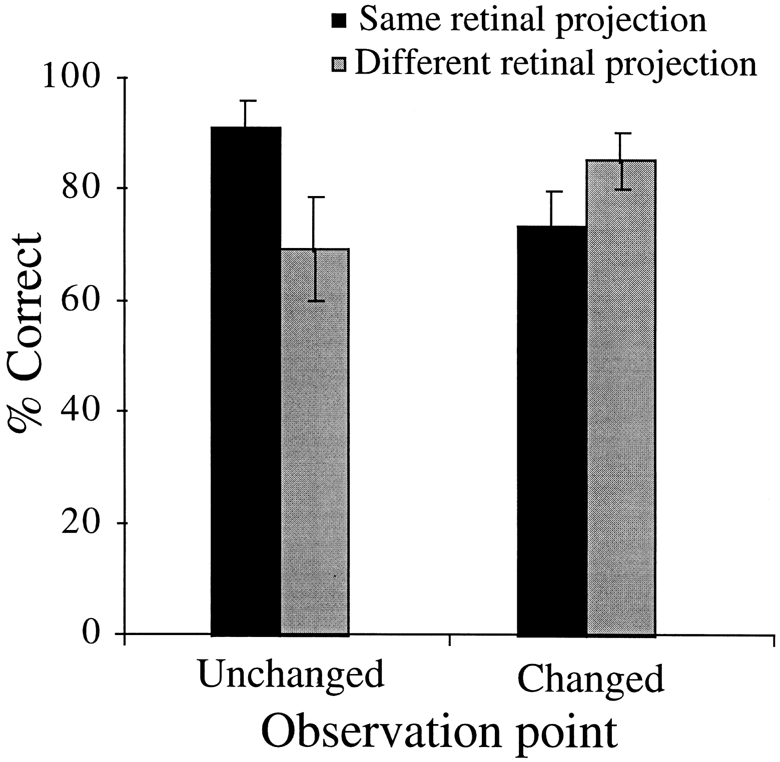

Fig. 2. Change detection accuracy in Experiment 1. Columns represent the percentage of correct responses. Error bars indicate 95% confidence intervals. When observers stayed at the same viewing position, they received a different retinal projection when the table rotated and the same retinal projection when the table did not rotate. When observers changed viewing position, they received the same retinal projection when the table rotated and a different retinal projection when it did not rotate.

on layout recognition (Simons and Wang, 1998) with two critical changes in the design. First, each subject completed all experimental conditions, thereby increasing the power of the comparisons. Second, the apparatus provided visual information (a pole attached to the table) for the timing and magnitude of the rotation whenever the table rotated. Even with the precise perceptual information for the change, observers at both observation points were most accurate when the table remained stationary. Strikingly, in the observer movement condition, observers were more accurate with a 40 degree view change than they were when they received the same view of the table at study and test. These findings suggest that in the real world we can recognize scenes and objects as we move, despite changes in the viewing angle. We apparently update our representations of objects as we move through our environment, and this updating process seems to disrupt our representation of the studied view of the scene. This ability to recognize scenes across views does not seem to generalize to objects rotating in front of stationary observers even when there is sufficient perceptual information for the magnitude of the object rotation; performance is significantly impaired by table rotations.[1]

One possible explanation for this important difference between viewpoint and orientation changes is that for viewpoint changes, observers actively control the magnitude of the change but for orientation changes, observers passively view the rotation. Active control over the change could facilitate performance by focusing attention on the magnitude of the change, thereby enhancing the precision of the representation. Or, active control might provide an efference copy of motor commands, thereby providing additional information for the magnitude of the view shift and allowing the representation to be updated.

Experiment 2 tested the role of active control by allowing the observers to physically cause a display rotation. If the difference between viewpoint and orientation changes is due to active control of the view change, then observers should be better when they actively rotate the table than when they passively view the rotation. If, on the other hand, the difference between the conditions results from something about the updating of viewer-centered representations as the observer moves and not from active control per se, observers should be no better when they cause orientation changes than when they passively view them; they do not move in either case. Experiment 2 examined the role of active control of the display rotation on the ability to detect object position changes across orientation changes.

7. Experiment 2

7.1. Method

The apparatus was the same as in Experiment 1. Eleven undergraduates participated in the study in exchange for $7 compensation. Unlike Experiment 1, observers remained at the same viewing position for all 40 trials of this experiment. On each trial, they viewed the array for 3 s (Study period) and then lowered the curtain. During the 7 s delay interval, the table rotated by 40 degrees. For half of the trials, the experimenter rotated the table (as in Experiment 1) and for the other half, the observer held the pole that was attached to the table and rotated the table themselves. Subjects then viewed the array again and indicated on a response sheet which object they thought had moved (Test Period). Thus, each subject received 40 trials with orientation changes. For half of the trials, they actively caused the orientation change. For the other half, they passively viewed the rotation. The order of the 40 trials was randomized for each subject.

---

[1] In this experiment, observers may have coded the position of objects with respect to the background of the room. Although this is a possible alternative interpretation of the results of this experiment, earlier research (described in Section 1) showed the same pattern of results with a darkened room and phosphorescent objects. Although the pragmatics of our current experimental setup preclude a replication of this glow-in-the-dark object study, the similarity of the results of the current experiment to our earlier one (Simons and Wang, 1998) is consistent with the notion that background information is relatively unimportant to the updating process.

7.2. Results and discussion

Observers were no more accurate when they actively caused the orientation change than when they passively viewed the change, $t(9) = 1.536$, $P = 0.1589$ (see Fig. 3). Performance in both conditions was comparable to performance for orientation changes in Experiment 1 (for active, $t(24) = 0.4018$, $P = 0.6914$; for passive, $t(24) = 0.3247$, $P = 0.7482$).

Contrary to the view that active control would facilitate the ability to update the representations across orientation changes, observers showed no signs of improvement when they actively controlled the display rotation. Although the motion of the rod may not provide the most vivid information about the rotation of the table (i.e. the mechanical coupling of the rod to the scene may be unfamiliar), this finding clearly demonstrates that performance impairments for orientation changes relative to viewpoint changes can not be attributed to a lack of information for the magnitude

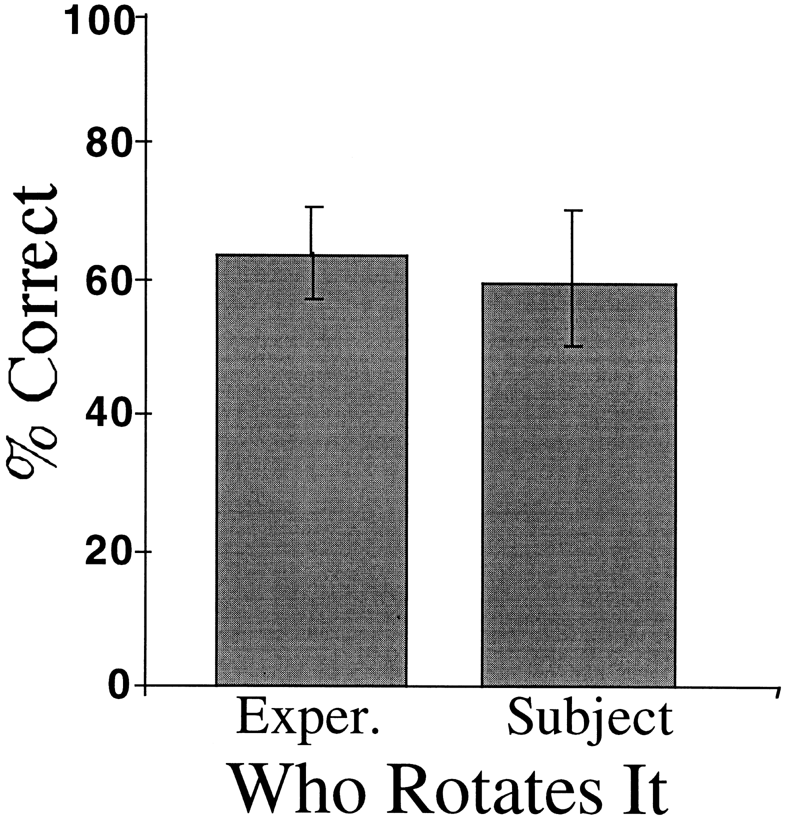

Fig. 3. Change detection accuracy in Experiment 2. Columns represent the percentage of correct responses. Error bars indicate 95% confidence intervals. When the experimenter rotated the table, observers passively experienced an orientation change. When the subject rotated the table, they actively controlled the change.

of the change or to the lack of active control over the change itself. Rather, the relative superiority of detection across viewpoint changes appears to result from a mechanism that is specific to view changes caused by observer movement. When observers are stationary, these mechanisms do not operate to update the representation even with reliable visual and motor feedback.

Although active control did not improve performance for orientation changes, it still may be critical for the ability to update representations across viewpoint changes. That is, superior performance in the viewpoint change condition may require both observer movement and active control of that movement. Experiment 3 examined the role of active control in the viewpoint change condition.

8. Experiment 3

In this experiment, observers sat on a wheeled chair and were rolled by an experimenter from the Study position to the Test position. If updating of the viewer-centered representation requires active control over the viewpoint change, observers should be less accurate when they are passively moved. By comparing

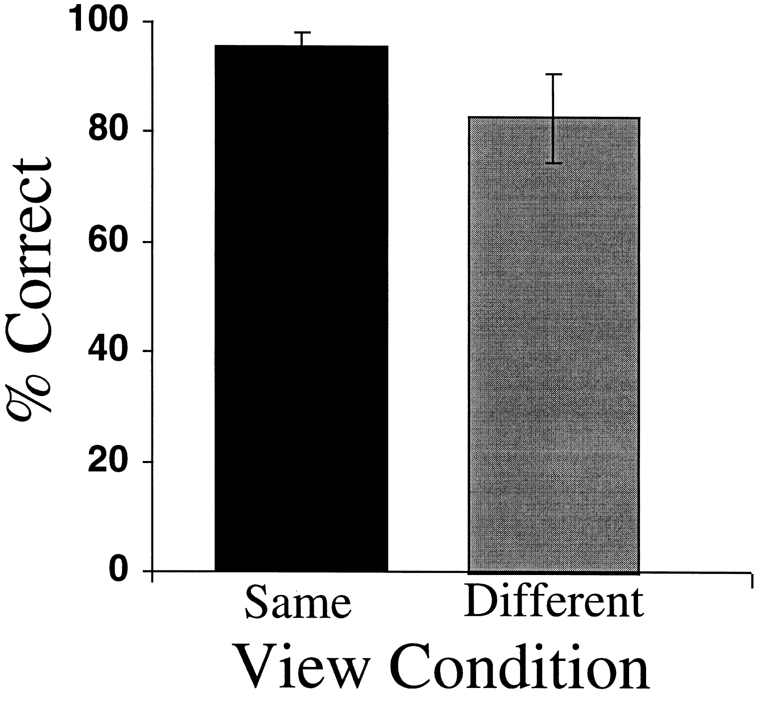

Fig. 4. Change detection accuracy in Experiment 3. Columns represent the percentage of correct responses. Error bars indicate 95% confidence intervals. In both conditions, the table did not rotate. Subjects received the same view when they remained at the initial viewing position and they received a different view when they were passively rolled to a different viewing position.

performance in this experiment to the corresponding active-movement condition in Experiment 1, we can access the effect of active movement on the updating process.

8.1. Method

The apparatus was the same as in Experiments 1 and 2. Ten undergraduates participated in the study in exchange for $7 compensation. On 20 trials observers received the same view at study and test (stationary observer, stationary table). On the remaining 20 trials, observers received a viewpoint change of 40 degrees (moving observer, stationary table). In the viewpoint change condition, observers sat in a wheeled chair and an experimenter passively rolled them from one window to the other during the delay interval. As in Experiment 2, the order of the 40 trials was randomized for each subject.

8.2. Results

As in Experiment 1, observers were slightly more accurate when they received the same view (from an unchanged viewing position) than when they received a different view from a new observation point, $t(9) = 3.822$, $P = 0.0041$ (see Fig. 4). Accuracy in the same-view, stationary-observer conditions of Experiments 1 and 3 did not differ, $t(24) = 1.537$, $P = 0.1374$. Accuracy when subjects were passively rolled to a different viewing position was comparable to the active walking condition of Experiment 1, $t(24) = 0.610$, $P = 0.5477$. Furthermore, even when subjects were passively moved, their performance was significantly better than that in all of the orientation change conditions of Experiments 1 and 2 (Exp. 1: $t(24) = 2.340$, $P = 0.0279$; Exp. 2: active $t(18) = 2.570$, $P = 0.0193$; passive $t(18) = 2.319$, $P = 0.0324$).

8.3. Discussion

When observers were passively rolled from the study position to the test position, their ability to detect a position change in the array was essentially as good as when they actively walked to the new viewing position. This finding suggests that active control of movement is not central to the process that allows the observers to recognize a scene from a novel position. Together with the results from Experiment 2, this Experiment demonstrates that active control over the change has little effect on the difference between orientation changes and viewpoint changes. Although orientation changes disrupt the ability to detect position changes in an array, viewpoint changes of an equivalent magnitude have a minimal effect on performance. The active-passive distinction does not account for this difference.

9. General discussion

When observers remain in the same position throughout a trial, they are better able to detect changes when they receive the same view at study and test. In striking contrast, when observers move to a novel viewing position during a trial, they detect changes more effectively when they receive the corresponding novel view than the studied view. That is, they are better able to detect changes when the orientation of the table is constant throughout a trial, even if that means they will experience a 40-degree view change from study to test (see also Simons and Wang, 1998). These findings suggest that real world scene recognition involves more than retinal images. Information about the ego-motion of the observer can greatly facilitate scene recognition from novel viewpoints. Thus, to fully understand how we perceive and act in the world we need to study these phenomena in actual environments and to consider the interaction of extra-retinal information with visual representations.

The experiments in this paper further demonstrate that neither active control of the array rotation nor additional visual information for the shift in view improves performance in the orientation change condition. Furthermore, passive transportation of the observer does not impair performance in the viewpoint change condition. These findings suggest that the difference in performance for array rotations and observer movement is not due to a lack of information. Instead, the difference appears to be in how the available information is used. This differential use of information by the updating system suggests that the updating process is specialized and that it readily incorporates information about viewer position changes but not other information indicating a view change.

The lack of an effect of active control in our experiments seems to contrast with earlier evidence that active movement plays a critical role in spatial cognition. However, the underlying mechanism and the task involved in our studies are different from those of earlier research. Active control may be particularly helpful for combining a sequence of views into a coherent, global spatial representation (e.g. Larish and Andersen, 1995; Peruch et al., 1995; Christou and Bülthoff, 1997) or for calibrating signals from different sensory systems (e.g. Held and Hein, 1958; Held and Bossom, 1961; Held and Freedman, 1963). However, in our studies, observers were not required to perform either of these tasks. Our evidence suggests that active control may not be critical when observers are asked to match a single novel view to a specific studied view.[2]

Although performance on viewpoint changes (mean = 85%, across all experiments) was clearly superior to performance on orientation changes (mean = 69%, across all experiments), performance for orientation changes was significantly above chance (20%). If representations were entirely view-dependent, why should we find better than chance performance for orientation changes? There are at least two possibilities. First, given that view-specific representations usually have a range of tolerance and our view change was just barely beyond this range (Bülthoff and Edelman, 1992), better than chance performance for orientation changes may be due

[2] Benson and Uzgiris (1985) showed that infants who actively crawled around a box were able to locate a target more accurately than when they were carried over, suggesting that at least during early development active movement may provide important cues for a process akin to updating. However, Acredolo et al. (1984) argued that it was not active movement per se, but visual tracking that caused the improvement.

to the relatively small angular distance between the studied and tested views (40 degrees). Alternatively, observers may encode some view-independent information about the array such as object-to-object relationships. If so, then performance might be better than chance whenever observers were able to use such representations. Informal observations and conversations with the subjects suggest that they did try to verbally encode some relationships among objects (e.g. the brush is to the left of the box). Other subjects tried to remember small clusters of objects and noticed when the cluster had changed. These crude view-independent (at least from the viewing positions used in this report) representations might allow them to detect changes on some subset of the trials, thereby leading to better-than-chance detection.

If subjects are using an effortful or verbal encoding strategy for orientation change trials but not for other types of trials, performance may be view invariant for some tasks and view specific in others, depending on which system the specific task requires. It is not clear that such effortful encoding would rely on the same systems used to detect changes in the same view condition or in the viewpoint change condition. For example, verbal interference might affect performance during orientation changes but not during viewpoint changes. Perhaps visual interference might influence performance on same view trials but not on orientation change trials. Future research could examine the nature of the representations underlying the orientation/viewpoint difference by systematically varying such distraction tasks.

Although performance across viewpoint changes was reliably better than that across orientation changes, accuracy was slightly reduced relative to the same view condition (stationary observer). This decrease could result from the intervening activities (moving from one observation position to another), from noise or inaccuracies in the updating process, or from specific errors in the observer’s position estimate as they move. The accuracy in position estimation may be particularly important when only partial information is available during observer locomotion (e.g. if observers were blindfolded, passively moved, spun around, etc.). Without complete information during locomotion, the updating mechanism itself may show some viewpoint dependence, albeit not to the same degree as for orientation changes. In particular, errors in observer position estimates may compound with increasing distance. Such possibilities could be examined by increasing the distance between viewing windows while preserving the magnitude of the view change (by moving the curtain further away from the table). If the errors result from observer position estimates, increasing the distance should increase the number of errors. Alternatively, if the errors primarily result from noise in the updating mechanism, error rates should not change provided that the magnitude of the view shift is comparable.

These experiments open many additional avenues for future research. For example, although passively-rolled subjects showed no deficit in updating their representations, it is unclear which of the remaining sources of information for the change (e.g. optical flow, vestibular feedback, proprioceptive feedback, etc.) are critical. Further, what would happen if observers translated toward or away from the array, causing a scale change rather than a viewing angle change? Given the parallels between these results and those from studies of imagined view changes, we have recently begun a series of studies directly comparing imagined and real view changes (both orientation and viewpoint). Perhaps the imagery tasks rely on the same updating mechanisms as the real-world tasks. Studies of the details of the recognition and updating process in real environments hold significant promise for the development of more realistic models of object recognition.

Acknowledgements

The authors contributed equally to this research and authorship order was determined arbitrarily. Thanks to Daniel Tristan and Chris Russell for help collecting the data and to M.J. Wraga for comments on an earlier version of the paper. Some of this research was presented at ARVO 1998.

References

Acredolo, C., Adams, A., Goodwyn, S.W., 1984. The role of self-produced movement and visual tracking in infant spatial orientation. Journal of Experimental Child Psychology 38, 312–327.

Amorim, M.A., Glasauer, S., Corpinot, K., Berthoz, A., 1997. Updating an object's orientation and location during nonvisual navigation: a comparison between two processing modes. Perception and Psychophysics 59, 404–418.

Amorim, M.A., Stucchi, N., 1997. Viewer- and object-centered mental explorations of an imagined environment are not equivalent. Cognitive Brain Research 5, 229–239.

Bartram, D.J., 1976. Levels of coding in picture-picture comparison tasks. Memory and Cognition 4, 593–602.

Benson, J.B., Uzgiris, I.C., 1985. Effect of self-initiated locomotion on infant search activity. Developmental Psychology 21, 923–931.

Berthoz, A., Israel, I., Francois, P.G., Grasso, R., Tsuzuku, T., 1995. Spatial memory of body linear displacement: what is being stored? Science 269, 95–98.

Biederman, I., Cooper, E.E., 1991. Object recognition and laterality null effects. Neuropsychologia 29, 685–694.

Biederman, I., Cooper, E.E., 1992. Size invariance in visual object priming. Journal of Experimental Psychology: Human Perception and Performance 18, 121–133.

Biederman, I., Gerhardstein, P.C., 1993. Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. Journal of Experimental Psychology: Human Perception and Performance 19, 1162–1182.

Bridgeman, B., Graziano, J.A., 1989. Effect of context and efference copy on visual straight ahead. Vision Research 29, 1729–1736.

Bridgeman, B., Stark, L., 1991. Ocular proprioception and efference copy in registering visual direction. Vision Research 31, 1903–1913.

Bridgeman, B., 1986. Multiple sources of outflow in processing spatial information. Acta-Psychologica 63, 35–48.

Bülthoff, H.H., Edelman, S., 1992. Psychophysical support for a two-dimensional view interpolation theory of object recognition. Proceedings of the National Academy of Sciences of the United States of America 89, 60–64.

Christou, C.G., Bülthoff, H.H., 1997. View-Direction Specificity in Scene Recognition After Active and Passive Learning. (Technical Report No. 53). Max Planck Institut für Biologische Kybernetik, Tübingen, Germany.

Cooper, E.E., Biederman, I., Hummel, J.E., 1992. Metric invariance in object recognition: a review and further evidence. Canadian Journal of Psychology 46, 191–214.

Diwadkar, V.A., McNamara, T.P., 1997. Viewpoint dependence in scene recognition. Psychological Science 8, 302–307.

Ellis, R., Allport, D.A., Humphreys, G.W., Collis, J., 1989. Varieties of object constancy. Quarterly Journal of Experimental Psychology 41A, 775–796.

Farrell, M.J., Robertson, I.H., 1998. Mental rotation and the automatic updating of body-centered spatial relationships. Journal of Experimental Psychology: Learning, Memory, and Cognition 24, 227–233.

Frier, H.J., Edwards, E., Smith, C., Neale, S., Collett, T.S., 1996. Magnetic compass cues and visual pattern learning in honeybees. Journal of Experimental Biology 199, 1353–1361.

Gale, N., Golledge, R.G., Pellegrino, J.W., Doherty, S., 1990. Acquisition and integration of route knowledge in an unfamiliar neighborhood. Journal of Environmental Psychology 10, 3–25.

Held, R., Bossom, J., 1961. Neonatal deprivation and adult rearrangement: Complementary techniques for analyzing plastic sensory-motor coordinations. Journal of Comparative and Physiological Psychology 54, 33–37.

Held, R., Freedman, S.J., 1963. Plasticity in human sensory-motor control. Science 142, 455–462.

Held, R., Hein, A.V., 1958. Adaptation of disarranged hand-eye coordination contingent upon re-afferent stimulation. Perceptual and Motor Skills 8, 87–90.

Huttenlocher, J., Presson, C.C., 1973. Mental rotation and the perspective problem. Cognitive Psychology 4, 277–299.

Huttenlocher, J., Presson, C.C., 1979. The coding and transformation of spatial information. Cognitive Psychology 11, 375–394.

Israel, I., Bronstein, A.M., Kanayama, R., Faldon, M., Gresty, M.A., 1996. Visual and vestibular factors influencing vestibular ‘navigation’. Experimental Brain Research 112, 411–419.

Kelso, J.A., 1977. Planning and efferent components in the coding of movement. Journal of Motor Behavior 9, 33–47.

Kirchner, W.H., Braun, U., 1994. Dancing honey bees indicate the location of food sources using path integration rather than cognitive maps. Animal Behaviour 48, 1437–1441.

Kourtzi, Z., Shiffrar, M., 1997. One-shot view invariance in a moving world. Psychological Science 8, 461–466.

Larish, J.F., Andersen, G.J., 1995. Active control in interrupted dynamic spatial orientation: the detection of orientation change. Perception and Psychophysics 57, 533–545.

Loomis, J.M., Klatzky, R.L., Golledge, R.G., Cicinelli, J.G., Pellegrino, J.W., Fry, P.A., 1993. Nonvisual navigation by blind and sighted: Assessment of path integration ability. Journal of Experimental Psychology: General 122, 73–91.

Nakatani, C., Pollatsek, A., Johnson, S., (1996). Mental rotation of scenes. Poster presented at the 37th Annual Meeting of the Psychonomic Society. Chicago, IL.

Peruch, P., Vercher, J.L., Gauthier, G.M., 1995. Acquisition of spatial knowledge through visual exploration of simulated environments. Ecological Psychology 7, 1–20.

Poag, C.K., Cohen, R., Weatherford, D.L., 1983. Spatial representations of young children: the role of self- versus adult-directed movement and viewing. Journal of Experimental Child Psychology 35, 172–179.

Poggio, T., Edelman, S., 1990. A network that learns to recognize 3D objects. Nature 343, 263–266.

Presson, C.C., 1982. Strategies in spatial reasoning. Journal of Experimental Psychology: Learning, Memory, and Cognition 8, 243–251.

Rieser, J.J., Garing, A.E., Young, M.F., 1994. Imagery, action, and young children’s spatial orientation: it’s not being there that counts, it’s what one has in mind. Child Development 65, 1262–1278.

Rock, I., Wheeler, D., Tudor, L., 1989. Can we imagine how objects look from other viewpoints? Cognitive Psychology 21, 185–210.

Ronacher, B., Wehner, R., 1995. Desert ants Cataglyphis fortis use self-induced optic flow to measure distances traveled. Journal of Comparative Physiology A: Sensory Neural and Behavioral Physiology 177, 21–27.

Shelton, A.L., McNamara, T.P., 1997. Multiple views of spatial memory. Psychonomic Bulletin and Review 4, 102–106.

Shepard, R.N., Cooper, L.A., 1982. Mental images and Their Transformations. MIT Press, Cambridge, MA.

Shepard, R.N., Metzler, J., 1971. Mental rotation of three-dimensional objects. Science 171, 701–703.

Simons, D.J., Wang, R.F., 1998. Perceiving real-world viewpoint changes. Psychological Science 9, 315–320.

Simons, D.J., 1996a. In sight, out of mind: when object representations fail. Psychological Science 7, 301–305.

Simons, D.J., 1996b. Accurate visual detection of layout changes requires a stable observer position. Investigative Ophthalmology and Visual Science 37(3), s519.

Spelke, E.S., Kestenbaum, R., Simons, D.J., Wein, D., 1995. Spatiotemporal continuity, smoothness of motion and object identity in infancy. British Journal of Developmental Psychology 13, 113–142.

Srinivasan, M.V., Zhang, S.W., Lehrer, M., Collett, T.S., 1996. Honeybee navigation en route to the goal: visual flight control and odometry. Journal of Experimental Biology 199, 237–244.

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 and Review 2, 55–82.

Tarr, M.J., Pinker, S., 1989. Mental rotation and orientation-dependence in shape recognition. Cognitive Psychology 21, 233–282.

Tarr, M.J., Bülthoff, H.H., Zabinski, M., Blanz, V., 1997. To what extent do unique parts influence recognition across changes in viewpoint? Psychological Science 8, 282–289.

Thinus-Blanc, C., Durup, M., Poucet, B., 1992. The spatial parameters encoded by hamsters during exploration a further study. Behavioural Processes 26, 43–57.

Ullman, S., Basri, R., 1991. Recognition by linear-combinations of models. IEEE Transactions on Pattern Analysis and Machine Intelligence 13, 992–1006.

Vetter, T., Poggio, T., 1994. Symmetric 3D objects are an easy case for 2D object recognition. Spatial Vision 8, 443–453.

Vetter, T., Hurlbert, A., Poggio, T., 1995. View-based models of 3D object recognition: invariance to imaging transformations. Cerebral Cortex 5, 261–269.

Wang, R.F., 1998. Imagination, action and spatial representations. Proceedings and Abstracts of the Annual Meeting of the Eastern Psychological Association 69, 91.

Wang, R.F., Simons, D.J., 1997. Layout change detection is differentially affected by display rotations and observer movements. Investigative Ophthalmology and Visual Science 38, s1009.

Wehner, R., Srinivasan, M.V., 1981. Searching behavior of desert ants, genus Cataglyphis (Formicidae, Hymenoptera). Journal of Comparative Physiology 142, 315–318.

Wraga, M., Creem, S.H., Proffitt, D.R., submitted. Updating scenes after imagined object- and viewer-rotations.

Xu, F., Carey, S., 1996. Infants' metaphysics: the case of numerical identity. Cognitive Psychology 30, 111–153.