VISUAL COGNITION, 2004, *11* (2/3), 341–370

Anticipating action in complex scenes

Ian M. Thornton

Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Amy E. Hayes

School of Sport, Health, and Exercise Sciences, University of Wales, Bangor, UK

In four experiments we explored the accuracy of memory for human action using displays with continuous motion. In Experiment 1, a desktop virtual environment was used to visually simulate ego-motion in depth, as would be experienced by a passenger in a car. Using a task very similar to that employed in typical studies of representational momentum we probed the accuracy of memory for an instantaneous point in space/time, finding a consistent bias for future locations. In Experiment 2, we used the same virtual environment to introduce a new ''interruption'' paradigm in which the sensitivity to displacements during a continuous event could be assessed. Thresholds for detecting displacements in ego-position in the direction of motion were significantly higher than those opposite the direction of motion. In Experiments 3 and 4 we extended previous work that has shown anticipation effects for frozen action photographs or isolated human figures by presenting observers with short video sequences of complex crowd scenes. In both experiments, memory for the stopping position of the video was shifted forward, consistent with representational momentum. Interestingly, when the video sequences were played in reverse, the magnitude of this forward bias was larger. Taken together, the results of all four experiments suggest that even when presented with complex, continuous motion, the visual system may sometimes try to anticipate the outcome of our own and others' actions.

Please address correspondence to: Ian M. Thornton, Max Planck Institute for Biological Cybernetics, Spemannstr. 38, 72076 Tübingen, Germany. Email: ian.thornton@tuebingen.mpg.de

The authors would like to thank Nils Aguilar for his help in filming and editing the digital video sequences used in Experiments 3 and 4. We would also like to thank Andries Hof, as well as two anonymous reviewers for their comments on earlier drafts of this paper. Further thanks go to our research assistants, Carrie Harp, Hong Tan, and Stephan Simon for helping to collect the data. The first two experiments were carried out at Cambridge Basic Research and we would like to acknowledge the support of Nissan North America Inc.

http://www.tandf.co.uk/journals/pp/13506285.html DOI:10.1080/13506280344000374

Representational momentum refers to the tendency for observers to *misremember* the stopping point of an event as being further forward in the direction of movement or change (Freyd & Finke, 1984). Such behaviour across a wide range of experimental situations—including frozen-action photographs (Freyd, 1983), implied rotation (Freyd & Finke, 1984; Munger, Solberg, Horrocks, & Preston, 1999), implied translation (Hayes & Freyd, 2002), implied disequilibrium (Freyd, Pantzer, & Cheng, 1988), and smooth, continuous motion (Hubbard & Bharucha, 1988; Verfaillie, de Troy, & Van Rensbergen, 1994)—is consistent with the idea that the visual system may sometimes try to *anticipate* the outcome of predictable events (Freyd, 1987, 1993; Hubbard, 1995b; Verfaillie & d'Ydewalle, 1991). While the precise mechanism(s) behind such behaviour is still a matter of some debate (e.g., Bertamini, 2002; Kerzel, 2000), its appearance across so many display types suggests that representational momentum[1] can be a useful probe into dynamic visual processing (Freyd, 1987; see Hubbard, 1995b, for a review; Thornton & Hubbard, 2002, for a recent collection of relevant papers).

The purpose of the current work was to extend our knowledge of the conditions under which representational momentum occurs by using complex simulated or videotaped scenes depicting human action. In Experiments 1 and 2, a desktop virtual environment was used to visually simulate movement in depth along a roadway from the perspective of a car driver or passenger. Our aim was to measure the accuracy of visual estimates of ego-position when the observer is also the (simulated) actor. In Experiments 3 and 4, we used digital video sequences of real-world crowd scenes to examine the accuracy with which the actions of others can be perceived and remembered.

The main characteristics of the displays just described—human action, scene complexity, and continuous motion—have all been individually explored in previous studies of representational momentum. For example, Freyd (1983), Futterweit and Beilin (1994), and Kourtzi and Kanwisher (2000) all used frozen action photographs of human figures to explore the processing of implied dynamics. Similarly, Karl Verfaillie and colleagues have focused specifically on the anticipation of human action using point-light displays (Johansson, 1975; Verfaillie et al., 1994) or computer animated figures (Verfaillie & Daems, 2002). In terms of scene complexity, a number of researchers have used multi-item

[1] The term ''representational momentum'' is used in two ways in the literature. In the first, and more limited sense, it is used to refer to a specific form of position displacement, possibly arising through the internalization of environmental invariants, where an analogy can be drawn with physical momentum. In this sense, representational momentum can be contrasted with other forms of displacement, such as representational gravity, friction, or centripetal force (Hubbard, 1995a, 1995b). In the second, and more general sense, the term is used as a shorthand way of referring to any form of displacement thought to involve anticipation in dynamic displays. It is this more general sense that we use throughout the current paper.

displays to explore various issues, such as multi-item tracking (Finke & Shyi, 1988), landmark attraction (Hubbard & Ruppel, 1999), and divided attention (Hayes & Freyd, 2002). Finally, the use of continuous motion rather than implied motion has been one of the hallmarks of Timothy Hubbard's work (e.g., Hubbard & Bharucha, 1988; Hubbard, 1995a, 1995b) and similar smooth motion displays have also been used in a range of other studies (e.g., Gray & Thornton, 2001; Verfaillie et al., 1994; although see Kerzel, 2000; Kerzel, Jordan, & Müsseler, 2001, for some issues arising from the use of simple smooth motion displays).

Clearly, in our everyday lives, these factors—action, complexity, and continuous motion—co-occur. One motivation for the current work was thus simply to know whether representational momentum would be observed when these three factors were brought together in experimental displays that were richer, and thereby that much closer to our everyday visual experience, than those used in previous studies. We expect that representational momentum should hold under these conditions; indeed, an explicit assumption has been that representational momentum reflects the mind's ability to represent the environment effectively in order to facilitate action in the real world (Freyd, 1987, 1993). Moreover, previous representational momentum studies have shown that a representation of physical invariants of the natural world, such as gravity and friction, are evoked even by rather impoverished, cartoon-like computer displays in which actual gravity and friction do not exist; this might imply that representational momentum is fundamentally rooted in capturing the dynamics of natural scenes (Hubbard, 1995a, 1995b). In the current paper we provide an empirical test of such assumptions. More specifically, we make a first attempt at establishing how the additional cognitive and perceptual processes involved in representing complex scenes interact with represented dynamics as measured by representational momentum.

Another, more general motivation for the current work relates to the relationship between vision and action. While it has long been accepted that a primary role of vision is to guide action (see Milner & Goodale, 1995 for review), more recent interest has also focused on the ways in which our actions, or even just our intended actions, help determine what we perceive (see Hommel, Müsseler, Aschersleben, & Prinz, 2001; Prinz & Hommel, 2002). In the current work, we do not directly manipulate the relationship between perception and action. Rather, our goal is first to establish that the perceptual effect of interest (i.e., representational momentum) can indeed be measured in complex, action-related scenes. Future work could then use such displays to more directly explore the link between perception and action in this context.

EXPERIMENT 1

The purpose of Experiment 1 was to examine whether the visual system can anticipate in complex environments. As we walk, drive, or fly through the world, our visual environment quickly changes in complex, though often predictable, ways. Indeed, some have argued that the information contained in such flow fields is rich enough to almost completely determine the visual structure of the world (Gibson, 1979). Despite the common occurrence of such dynamic patterns and the obvious utility of being able to predict change in complex environments, little work to date has directly examined anticipation in this context. Previous research has established that when viewing other objects moving in depth, forward memory shifts for object position do occur (e.g., Hayes, Sacher, Thornton, Sereno, & Freyd, 1996; Hubbard, 1996; Nagai, Kazai, & Yagi, 2002). Anticipation of self-motion, however, has not been investigated (although see DeLucia, 2001, for some recent experiments in this direction). The goal of the present study was thus to directly assess the accuracy of visual memory for ego-position during movement in depth within a scene using a representational momentum paradigm.

To do this, we used a desktop virtual environment to simulate a drive through a novel scene. At a random point during the smooth motion through the scene, a 250 ms blank interval was inserted. The observers' task was to remember their position in the scene immediately prior to this blank interval. After 250 ms, the animation was restarted; on some trials the starting point was identical to the stopping point and on others it could vary, as if the observer had been shifted either forward or backward during the blank interval. Observers were asked to judge whether their new position was the same as their remembered stopping point, or was in any way different.

Our interest was in how accurately observers could extract an impression of instantaneous ego-position from such a complex, dynamic flow field. More specifically, based on previous representational momentum findings, we predicted that sensitivity to forward changes in position would be reduced compared to backward changes in position. This would be consistent with observers anticipating their position in space.

Methods

Participants. Twelve observers from the Boston/Cambridge community were paid for participation in this study. All observers had normal or corrected to normal vision and were naive as to the purpose of the research until after the experimental session ended.

Equipment. A virtual environment was created and presented on a Silicon Graphics workstation connected to a standard 20-inch colour monitor with a resolution of 1280 (horizontal) $\times$ 1024 (vertical) pixels and refresh rate of 75 Hz. Observers sat 65 cm from the monitor, so the visible portion of the display subtended 24° vertically and 41° horizontally.

Stimuli and design. The virtual environment contained a straight, single-lane, grey road receding in depth over a green, textured ground plane. The roadway had a simulated length of 200 m in total, and this was divided into three sections, in which only the central 100 m contained landmarks. The viewing parameters were set so that the entire 200 m landscape was visible at the start of the trial. The overall impression was that of a viewer from a car that approaches a small village or hamlet, drives through the hamlet, and then continues along the open road.

Within the central 100 m a series of landmark objects were randomly placed on either side of the roadway. Landmarks were used to provide a more detailed scene context and to improve the sense of motion in depth. The landmarks consisted of trees, benches, office buildings, houses, and signposts. Landmarks differed from each other in colour, size, and shape. The largest object was an office building, which had a simulated size of 15 × 10 × 5 m, the smallest was a bench which measured 5 × 2 × 1 m.

A custom routine was created that randomly selected landmarks on each trial and distributed them in pseudorandom locations along either side of the central portion of the roadway. More specifically, the edges of each side of the road were divided into a series of 20 m ‘‘bins’’. On any trial, a bin could be empty or could contain a random landmark. The precise position of the landmark was randomly varied within the central 10 m of the bin. The distance between the road edge and the landmark was also randomly varied between 3 and 5 m. The purpose of varying the density and layout of the landmarks on each trial was to create the impression that a novel scene was being presented on each trial. This was done to avoid the situation where observers could focus on one given point or object as a reference frame with which to judge their position from trial to trial.

The route was presented as a continuous drive along the road, with a viewing height of 1.5 m above the ground. On any given trial the speed throughout the animation was held constant at 16, 18, or 20 m/s. These speeds correspond to approximately 58, 65, and 72 km/h, respectively. Speed was varied to increase the variety of the task and also to explore whether previously reported velocity effects for representational momentum (e.g., Finke, Freyd, & Shyi, 1986) would be replicated here.

The blank interval was randomly positioned on a trial-by-trial basis and could fall anywhere within the central 100 m of the route. After the 250 ms blank interval, on true/same trials, the animation resumed from exactly the same location that it had stopped. This occurred on 1/11 of the trials. On the remaining probe trials, the animation could start at 3, 6, 9, 12, or 15 m ahead or behind of the true point of interruption. Thus, there were five forward probe locations and five backward probe locations that appeared with equal probability.

The parameters of the display were coded in units of virtual meters, and the metrics associated with all scene features were scaled to be consistent with actual navigation. Accordingly, throughout a trial the degrees of visual angle subtended by each visible object changed consistently over time. For example, a building having the dimensions 10 × 8 × 5 meters would first appear in the far distance as a very simple polygon structure with a height of, say, 10 pixels. At the viewing distance of 65 cm, the building would subtend approximately 0.33° visual angle. As the simulated car approached the building—assuming it was quite close to the edge of the road—the walls of the building might take up the entire vertical extent of the screen, subtending almost 24° visual angle. The precise layout of the scene at the moment when the animation was interrupted and then later probed was not recorded. However, over multiple trials the random stopping point would give rise to a great variety of landmark configurations. While the particular positions of landmarks may influence memory for the scene (Hubbard & Ruppel, 1999), we here investigate only the overall bias that occurs over a wide variety of landmark configurations.

Task. The observers in this task had no control over the simulated vehicle, either in terms of its speed or position on the roadway. Thus, their role was more a passenger in a moving vehicle, rather than the driver. Their task was to passively view the initial animation sequence and to try and remember their precise ego-position in depth immediately before the screen was blanked. This remembered position was then to be explicitly compared to the position that appeared immediately after the blank interval. Observers were instructed to respond ''same'' if they judged that their position had not changed in anyway during the blank. They were to respond ''different'' if they perceived any change, either forward or backward, during the blank. Responses were made via two keys on a standard keyboard.

Procedure. Observers were first made familiar with the general nature of the display, the task, and the method of responding through a series of demonstration trials. Once they were comfortable with the experimental environment they completed a block of 20 practice trials, which were randomly selected from the full experimental design. Each participant then completed 198 experimental trials consisting of 3 speeds × 11 probe positions × 6 repetitions. These factors were randomly intermixed and trial order was determined separately for each observer. A break was provided after 66 and 132 trials.

Results

Figure 1 shows the percentage of ''same'' responses as a function of probe distance from the true/same stopping point, collapsed across observers and velocity. The peak of this function is shifted forward, suggesting a reduced sensitivity to probe trials that were ahead in the direction of motion rather than opposite to the direction of motion. To estimate the magnitude of this forward

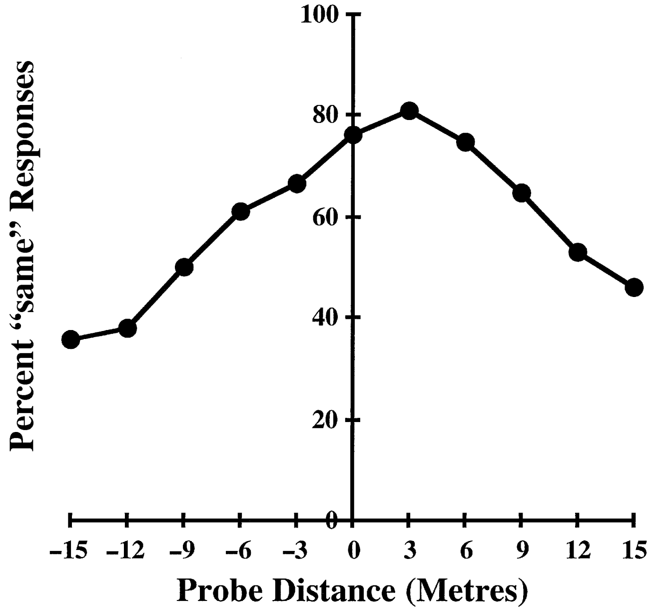

Figure 1. Percentage ''same'' responses as a function of probe distance for Experiment 1. Probe distance is measured in metres of simulated distance travelled from the true/same stopping point.

bias, the central tendency of the distribution of same responses was calculated using a weighted mean (Faust, 1990; Hayes, 1997). The weighted mean for each observer is calculated by multiplying the proportion of same responses at a given probe position by that probe's distance from the true/same probe (i.e., 0). These products are then added and divided by the total number of same responses to yield a weighted mean. For the distribution in Figure 1, the average weighted mean was 0.97 m (SE = .30), a value that was reliably greater than zero, $t(11) = 3.34$, $p < .01$.

To assess the impact of speed of movement on the overall pattern of results, weighted means were calculated separately for the 16, 18, and 20 m/s velocities. These means, averaged across observers, were 0.69 (0.27), 1.28 (0.52), and 0.99 (0.35) m respectively (standard errors in parentheses). These means were all reliably greater than zero, $t(11) > 2.5$, $p < .0125$ in one-tailed tests, but there were no significant differences between them, $F(2, 22) = 0.94$, $MSE = 1.1$, n.s.

Discussion

The results of Experiment 1 indicate that estimates of instantaneous ego-position at the point of interruption were biased in the direction of motion. That is, even with a more complex display the data clearly show the typically forward memory shift associated with representational momentum. These results suggest that, as with simple object motion, we may also tend to anticipate our own movements through space.

While the current task closely approximates previous representational momentum studies in many respects, one clear difference is the nature of the probe. Typically, when the probe method is used to assess memory displacements (e.g., Freyd & Johnson, 1987), as opposed to a direct method of localization, such as pointing with the mouse (e.g., Hubbard & Bharucha, 1988), a static item is presented until a response is made. In the current study, the observer is asked to compare a remembered position to the onset position of a new sequence of animation. We chose this technique because in pilot testing with a static probe the displays gave rise to a very strong and very salient negative motion aftereffect. That is, when a static probe was shown, observers experienced a strong illusory impression of motion in the opposite direction (i.e., a contraction of the display, as if the observer were moving backward). Such motion aftereffects (MAEs) have been studied for many years (e.g., the waterfall illusion) and can be experienced with a wide range of display types (see Mather, Verstraten, & Anstis, 1998, for a review). Indeed a number of studies from the same laboratory as the current experiment have used very similar virtual environments (albeit with much longer adaptation times, e.g., 20 min) to explore the implications for such effects on driving decisions (Gray & Regan, 2000).

The use of a dynamic probe sequence greatly reduces, if not completely eliminates, the *subjective impression* of an MAE. That is not to say such a dynamic probe removes the influence of any adaptation that may have occurred. For example, Gray and Regan (2000) showed that exposure to motion in depth in a driving simulator could consistently influence estimates of speed in a subsequent task many minutes after all traces of a measurable MAE were gone. Bertamini (2002) has recently pointed out that such adaptation to motion in the inducing direction during representational momentum tasks might well be expected to reduce sensitivity to subsequent probes that differ from the true stopping point by displacements in that direction, thus influencing, if not fully accounting for, observed biases. Clearly, such factors may be operating in the current studies (although see Kerzel, 2003, for some conflicting findings).

From a practical standpoint, the use of dynamic probes allowed us to remove one source of subjective difficulty for observers (i.e., the experience of the MAE interfering with same/different judgements). However, such probes may also have influenced performance in at least two other respects. First, the complexity of the continued motion sequence after the blank interval may itself have interfered with same/different judgements by masking memory for the to-be-remembered point of interruption. For example, presenting an unrelated pattern or object *during* the retention interval has been shown to alter patterns of forward displacement in simple displays (Thornton, DiGirolamo, Hayes, & Freyd, 1996). Such masking is thought to occur because information about the to-be-remembered item is disrupted or replaced in some way (Di Lollo, Enns, & Rensink, 2000). Some form of substitution masking could occur in our displays if the rate of change in the dynamic probe introduces new or unrelated items. Kirschfeld and Kammer (1999) have also suggested that—at least with simple motion displays—subsequent positions of a target item mask previous positions through a process of metacontrast. Such a process operating in our displays could reduce the perceptual availability of the to-be-remembered point of interruption.

A second way in which the dynamic probes may influence performance concerns the degree to which observers are able to veridically perceive the onset position of the probe animation for comparison with the to-be-remembered point of interruption. There is considerable evidence from studies using simple object motion to suggest that the onset of a moving target can often be mislocalized, either in the direction of motion (i.e., the Fröhlich effect; Fröhlich, 1923; Müsseler & Aschersleben, 1998) or opposite the direction of motion (i.e., the onset repulsion effect; Actis-Grosso, Stucchi, & Vacario, 1996; Hubbard & Motes, 2002; Kerzel, 2002a; Thornton, 2002b). Understanding when an observed onset effect will be in the direction of motion versus opposite the direction of motion is the focus of much ongoing research. It seems that a number of task parameters, such as the predictability of target motion (Müsseler & Kerzel, 2003; Thornton, 2002b), the response mode (Kerzel, 2002a) and even the surrounding context (Hubbard & Motes, 2003; Thornton, 2002a), all help to determine the nature of the observed shift.

In the current experiment, a Fröhlich effect would oppose the representational momentum effect, contributing a memory bias in the negative direction. The onset repulsion effect would produce a forward memory bias, consistent with the representational momentum effect. It is not immediately clear how, if at all, either of these onset effects could contribute to the current pattern of observed data, but clearly further use of a dynamic probe task would need to take these factors into account.

An unexpected finding in the current study was that the manipulation of travelling speed did not affect the size of the memory shift. Based on previous representational momentum studies, one would predict an increase in the magnitude of the forward shift with increasing implied or actual motion. Indeed, it was the finding of an almost linear relationship between velocity and memory shift in earlier studies that helped support a direct relationship between physical and ''internal'' momentum (Finke et al., 1986). It is unclear why we did not observe a velocity effect here, although several of the novel aspects of our displays, such as the use of motion in depth, ego-motion, complexity, etc., could have contributed.

In conclusion, our data support the notion that we may tend to anticipate our own movements through space. But the conclusion that representational momentum occurs for ego-motion requires a qualifying remark. Although we have greatly increased the richness of the display compared to most studies of representational momentum, a consideration of the methods section will suggest that observers are very unlikely to believe they are actually moving through space. Thus, at best, we have been able to partially simulate some aspects of the visual component of ego-motion. Further studies using richer virtual environments or in which the observer actually moves through the scene might provide stronger evidence concerning memory for ego-position.

EXPERIMENT 2

The purpose of the previous experiment was to explore memory for ego-motion with a task that was as similar as possible to the standard representational momentum paradigm. However, in this more complex domain, the representational momentum task proved to be rather difficult for observers. Even with displacements as great as 15 m forward or backward, many observers in Experiment 1 were still unable to detect a displacement on large numbers of trials. That is, the tails of the distribution shown in Figure 1 remain quite close to chance (50%). The purpose of the current experiment was to introduce a new type of ''interruption'' paradigm, which we believe to be a more natural and less demanding way of assessing sensitivity to change in an ongoing dynamic stream.

The basic idea is as follows: Imagine that you are driving along at a constant speed during a heavy rainstorm. Passing traffic in the other lane periodically splashes so much water onto your windscreen that your field of vision is temporarily blocked. When vision returns, the spatial position of your car has been updated in accordance with the speed of the car and the duration of the visual occlusion.

So far, we have described what would happen in the real world and this scenario is also our default ''true/same'' experimental case. Generally, we seem quite capable of coping with the spatial updating issues associated with such visual occlusions. We contrast this ''true/same'' situation with one in which the relationship between space and time is experimentally manipulated. More precisely, on ''different'' trials, we systematically vary the spatial position of the point following the central occlusion so that the world that reappears corresponds either to a shift ahead or a shift behind the true/same location.

Our interest is to measure the sensitivity to such changes, and in particular, to assess whether there are any differences between forward and backward shifts. Rather than using the fixed probe method of Experiment 1, we used an adaptive staircase method, where the size of the forward and backward shifts were reduced to the point where they could be detected on approximately 71% of trials. Of interest was whether there would be an asymmetry in the sensitivity thresholds associated with forward versus backward shifts. The results of

Experiment 1, and previous studies of representational momentum would predict higher thresholds (i.e., lower sensitivity) for forward displacements, consistent with some form of perceptual anticipation of action.

Methods

Participants. Six experienced psychophysical observers took part in this experiment. All were members of either the Nissan Cambridge Basic Research lab or the University of California, Davis, Center for Neuroscience. Two of the observers (IMT, AEH) were authors; the other four were naive as to the purpose of the research.

Equipment and stimuli. The basic experimental environment was identical to that used in Experiment 1. On each trial, a new road segment was generated with random placement of landmark objects to the left and to the right within the central 100 m of the route. Speed of movement was constant and in this experiment was always 18 m/s. Rather than the single interruption or occlusion used in Experiment 1, here three brief (250 ms) occlusions were introduced, each separated by 250 ms of normal animation. Following Occlusions 1 and 3, the animation resumed at a point 4.5 m ahead of the disappearance point, consistent with continued, smooth motion. After Occlusion 2, the spatial position of the animation either resumed in the appropriate place, 4.5 m ahead (same trials) or was shifted forward or backward according to an adaptive staircase procedure (different trials). The staircase procedure is described in more detail below, but essentially the size of the introduced shift was manipulated independently for each subject so that they correctly detected a difference on 71% of trials.

Note that the spatial position of the onset of Occlusion 3 was always relative to the shift introduced in Occlusion 2. That is, after Occlusion 2, there was 4.5 m of normal animation followed by Occlusion 3. Thus, the only spatiotemporal perturbation occurred following Occlusion 2. The location of the initial occlusion was randomly located within the central 50 m of the entire route. This constraint (in Experiment 1 the interruption could occur anywhere within the central 100 m) was imposed to ensure that all three occlusions would be completed within the region of the roadway that contained landmarks.

Task and procedure. The observers' task was to detect an ''unusual'' interruption in the animation sequence. Note that on each trial, there would always be three occlusion interruptions. Observers had to try and discriminate the default ''true/same'' case, where the spatial location was correctly updated during the occlusions, from a situation where their position shifted unusually—either forward or backward—during the middle occlusion. To fully familiarize observers with the default case, each session began with a series of demonstration trials in which only the default case was shown. When observers were comfortable with the default case, a series of trials were shown in which a large displacement (25 m) was added to the middle occlusion. Thus, they were familiarized with an extreme example of an unusual displacement. When observers clearly understood the nature of the task, the main experimental session began.

Design and data analysis. Data were collected from each observer in a single session. Two adaptive staircases, one initially forward and one initially backward, were interleaved with each other and with a series of default true/same trials. For each trial, one random process was used to select between a default and a staircase trial. When a staircase trial was selected, a second random process was used to select between forward and backward staircases.

Responses to default trials were not analysed. A standard 2-up, 1-down transformed response method (e.g., Wetherill & Levitt, 1965) was used to adapt the difficulty of the task. For both forward and backward staircases, the initial displacement was set to 25 m, a value that pilot testing indicated should be easily detected by all observers. A correct response in the current context is responding ''different'' to a staircase trial; an error is responding ''same''. Following two correct responses the magnitude of the displacement was reduced (i.e., the discrimination was made more difficult) and following a single error the displacement was increased (i.e., the discrimination was made easier). Thus, at any moment in time the general trend of a staircase could be of either increasing or decreasing difficulty. A reversal in such a trend typically occurs close to the discrimination threshold and in the current study a staircase ''terminated''—was no longer available for random selection—when 18 such reversal points had occurred. The entire session ended when both adaptive processes had terminated. The threshold for each observer and each staircase was estimated by averaging across the final eight reversal points. This threshold approximates an accuracy level of 71% correct.

Results

On average observers needed 142 (SE = 10) trials to complete this experiment. Excluding ''same'' trials, the forward staircase (M = 43, SE = 3) terminated a little faster than the backward staircase (M = 50, SE = 4), with this difference reaching marginal significance, $t(5) = 2.3$, $p = .06$. Figure 2 summarizes the thresholds for each observer, both for forward and backward displacements. It is clear that for all observers, detection of a forward displacement (M = 14.5 m; SE = 2) is more difficult than for backward displacement (M = 9.4 m; SE = 1.5). In general, it seems that this difference is more pronounced for the four naive observers. In any event, across all six observers this difference between forward and backward thresholds was reliable, $t(5) = 4.5$, $p < .01$.

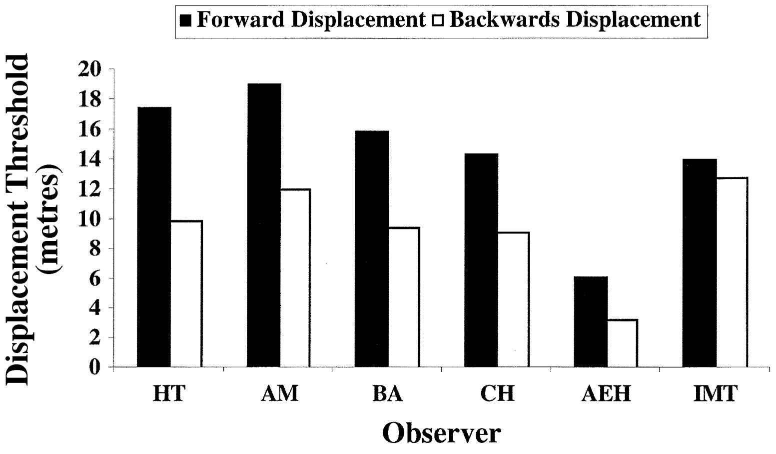

Figure 2. Thresholds for detecting forward or backward displacements in position along simulated roadway in Experiment 2.

Discussion

Using the same basic display environment, but a different, possibly more natural, method of probing, we again found a pattern of results consistent with anticipation in the direction of simulated ego-motion. That is, all observers showed better sensitivity for the detection of backward displacements (i.e., against the flow of motion) compared to forward displacements.

To the extent that these data can be compared to the results from Experiment 1, the magnitude of bias appears to be approximately the same, but overall accuracy is improved. Averaging across observers in Experiment 2, a negative probe of 9.4 m and a positive probe of 14.5 m each produce error rates of 29% (the target error rate set by the staircase). The midpoint of these probe positions, 2.6 m, gives an estimate of the average forward bias in Experiment 2. By comparison, in Experiment 1 the peak of the distribution of same responses occurs at the probe position of 3 m; the shift estimates in the two experiments thus appear to be of similar magnitude. Error rates, however, differ between the two experiments. The probe positions of –9.4 m and +14.5 m, which in Experiment 2 correspond to 29% error rates, in Experiment 1 corresponded to higher error rates of just over 40% (see Figure 1). Thus it appears that observers do in fact find the novel task used in Experiment 2 less difficult. Whether this increased accuracy in Experiment 2 is due to the more naturalistic probe task or due to the staircase method of trial presentation remains to be investigated. Nevertheless, as far as can be estimated, the magnitude of the memory bias appears to be roughly comparable to that revealed by Experiment 1.

While with this initial experiment we cannot rule out several other interpretations as to the source of this asymmetry—for example, some form of simultaneous masking in the direction of motion (e.g., Kirschfeld & Kammer, 1999)—our goal here was mainly methodological, that is, to develop a task that was less demanding than trying to explicitly retain an instantaneous impression of position within a continuing animation sequence. While we believe we have gone some way in achieving this goal, it is clear that further studies will be required to more fully explore the potential of this new ''interruption'' paradigm. These initial results at least demonstrate that such a task can be applied to the measurement of ego-position.

EXPERIMENT 3

In Experiments 1 and 2 we explored whether the visually simulated action of an observer—i.e., driving into a scene—gave rise to anticipatory errors in remembered ego-position. In both experiments the pattern of errors was consistent with such anticipation, although as just discussed, we believe that such results will require further investigation. In the current experiment, we return to the more common experimental scenario of a static observer, but continue with the theme of action by exploring memory for human crowd scenes. To do this we filmed a number of naturally occurring crowd scenes in and around a major German city. Figure 3 shows a representative still frame from each of the four clips used in Experiment 3. The common feature is that a crowd of human figures (i.e., more than 10 people) have been filmed with a static camera and are thus translating relative to the observer. More details concerning the content of each video sequence is given in the methods section below.

On each trial of this experiment, a short movie sequence was randomly selected and presented at full frame rate, that is, in real time with smooth motion. The screen was then blanked for 250 ms, and a static probe image was presented until response. This probe image could either correspond exactly to the stopping point of the inducing sequence or could systematically vary, forward or backward, around that point. The observers' task was to detect any difference between the stopping point of the inducing clip and the probe image. Of interest is the distribution of ''same'' responses as a function of probe position, with representational momentum characterized by a tendency to endorse those probes that fall ahead, i.e., in the direction of motion or change.

Our main motivation in designing this experiment was to assess whether representational momentum could be observed using films of continuous, complex human action. While anticipation of human action has been reported in a number of other studies, these have either used frozen action photographs (e.g., Freyd, 1983) or isolated computer-animated figures (e.g., Daems & Verfaillie, 1999). If we can demonstrate such anticipation effects using video clips of natural actions—and the more general method of probing described below—this might open the door for future work which more directly manipulates the

Figure 3. Representative frames from the four crowd-scene films used in Experiment 3. Top left: Clip 1: High school entrance; top right: Clip 2: Town square, left view; bottom right: Clip 3: Department store, interior; bottom left: Clip 4: Railway station, left view. Example movie clips for each sequence (and those for Experiment 4) are available online. See text for details.

relationship between perception and action, for example, by comparing the perception of ones own videotaped actions versus those of another (cf. Knoblich & Prinz, 2001).

As with Experiments 1 and 2, the move to more complex displays also involved a number of methodological changes which differentiate the current work from previous studies. Primary among these was the use of probe images which shared only a constant temporal relationship to the true/same stopping point and could vary quite dramatically in content (i.e., magnitude of difference from true/same) from trial to trial. We return to consider these implications of such methodological changes in the General Discussion.

Methods

Participants. Twelve observers from the Tübingen community were paid for participation in this study. All observers had normal or corrected to normal vision and were naive as to the purpose of the research until after the experimental session ended.

Equipment. The crowd scenes were filmed using a standard digital video camera with a sample rate of 25 frames/s. Each of the four basic sequences was trimmed to a duration of 10 s and exported as a sequence of 250 uncompressed image files. Custom presentation software was used to present the images as a continuous movie on a standard colour monitor using a screen resolution of 800 × 600 pixels and a refresh rate of 120 Hz. From a fixed viewing distance of 60 cm, the total display area subtended approximately 12 × 10° visual angle.

Stimuli. The four crowd scenes used in this experiment were all filmed in and around Stuttgart, Germany during March–April 2001. Figure 3 shows a characteristic still frame from each of the four clips. A 2 s image-compressed example movie of each scene can be viewed and/or downloaded from http://www.kyb.tuebingen.mpg.de/links/crowds.html. While the image quality of these clips is somewhat reduced from that used in the actual experiments, the frame rate is correct, so a clear indication of the magnitude of difference between successive probe images can be obtained by frame-advancing through the movies.

All scenes were filmed using natural, indoor or outdoor lighting, the camera was static and the subjects captured in the clips were typically unaware that they were being filmed. Scenes were selected so that crowd density and speed of movement in the scene varied across clips. Next, we briefly describe the set-up and content of each of the clips in more detail:

- *Clip 1 (high school entrance).* This clip shows staff and pupils leaving a local high school. The crowd density is relatively low, between 10 and 15 people in a given frame, and the pace of walking is relatively slow, estimated from the video sequence to be around 30 strides (i.e., left foot forward to left foot forward) per minute.
- *Clip 2 (town square, left view).* This clip shows a panoramic view of a large city square. Crowd density is high, with between 50 and 70 people visible in a given frame and walking speed is also typically high, at around 50 strides per minute.
- *Clip 3 (department store, interior).* This clip is an interior shot of shoppers in a local department store. Between 10 and 20 shoppers move relatively slowly along two aisles, sometimes stopping to look at articles in the central display areas. In the foreground, between 1 and 5 shoppers move along a walkway that lies perpendicular to the two shopping aisles.
- *Clip 4 (railway station, left view).* This clip shows travellers who have recently exited a train (visible at the left edge of the shot) moving along the platform. There are between 4 and 6 passengers in full-body close-up, with between 10 and 20 other passengers further down the platform. Walking speed, estimated from the clip, was fairly uniform and relatively fast at around 60 strides per minute.

Design. In the current experiment, a blocked design was used so that all trials relating to a single crowd scene were completed before moving on to the next. The order of the clips was randomly determined on an observer-by-observer basis. During an experimental trial, 10 consecutive frames (400 ms) were randomly selected from within the full 10 s clip. The probe frame was also randomly selected and could be either identical to the stopping point (true/same) or either 4, 8, or 12 frames ahead of or behind the stopping point. As the digital camera sampled at 25 frames/s, these ''different'' probe locations correspond to 160, 320, or 480 ms forward or backward in time. Given the nature of the stimuli and the method of sequence selection, these temporal offsets will be used to quantify the nature of responses irrespective of the image content. The blank retention interval between the end of the clip and the presentation of the probe was fixed at 250 ms.

Task. The task of the observers was to watch the brief video clip and try to remember exactly where the action stopped immediately before the blank interval. When the static probe image appeared, observers were instructed to respond ''same'' if they believed they were looking at exactly the same frame as had last been shown or ''different'' if the still image was in any way altered. Responses were made via two keys on a standard computer keyboard.

Procedure. At the beginning of the experimental session observers were shown a diagram depicting a sequence of video frames and demonstrating the relationship that the various forms of probe image could have to the stopping point. They were then given as many demonstration trials as they needed to become familiar with the basic task and method of responding. The sequence used for demonstration depended on their random block order. Before each of the four experimental blocks a training block of 20 trials was given, randomly sampled from the full design, to familiarize them with the content of the new video sequence. The full experimental block was then completed. This consisted of 70 trials (10 repetitions $\times$ 7 probe locations). A brief pause was given between successive blocks.

Results

Figure 4a shows the proportion of same responses collapsed across observers and video clips. The distribution is clearly biased with a central tendency—estimated by calculating the average weighted mean across observers—of approximately 40 ms, a figure consistently greater than zero, $t(11) = 5.9$, $p < .0001$.

To assess whether this pattern of results occurred for all of the video sequences, separate weighted means were calculated for each clip. As can be seen in Figure 4b, for three of the four clips the level of forward bias was fairly

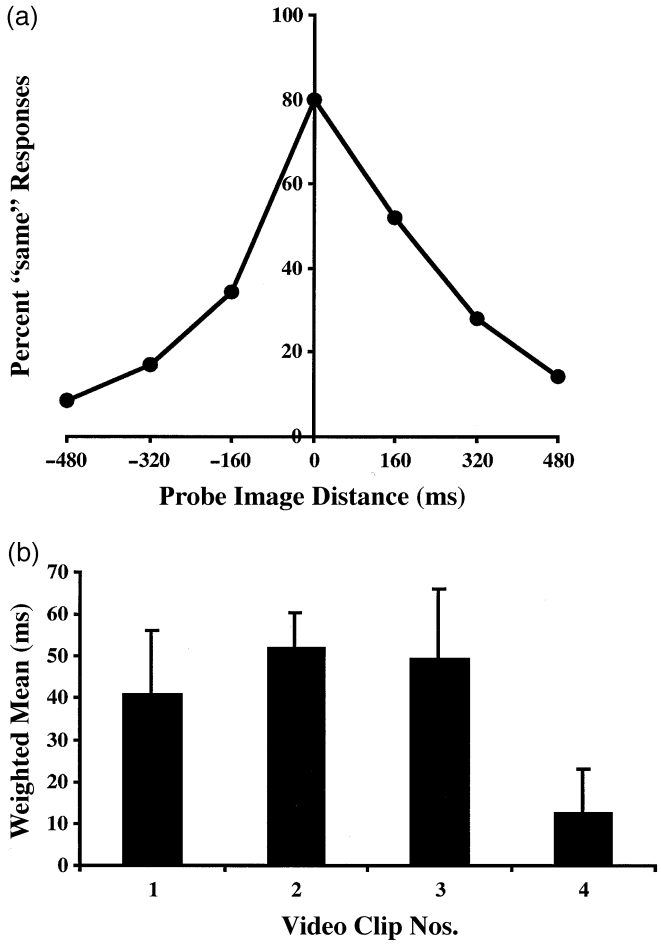

Figure 4. (a) Percentage ''same'' responses as a function of probe distance in Experiment 3. Probe distance is measured in units of time along the video clip from the actual stopping point. (b) Estimated memory shifts for the four video clips in Experiment 3.

constant, falling between 40 and 50 ms. However, for the scene filmed at the train station (Clip 4), responses were much less biased ($M = 12$ ms). Despite this drop in bias for Clip 4, a one-way repeated measures ANOVA failed to find any significant difference among the four clips, $F(3, 33) = 2.2$, $MSE = 1777$, $p = .1$. Nevertheless, comparing each weighted mean directly against zero showed that while Clips 1–3 were reliably different from zero, $t(11) > 2.5$ and $p < .0125$ in one-tailed tests, the mean for Clip 4 was not, $t(11) = 1.3$, n.s.

Discussion

The overall pattern of results in Experiment 3 show that memory for the stopping point of a short video sequence is biased in the direction of motion. As with the results of Experiment 1, the current findings indicate that representational momentum can be found in complex displays involving continuous motion. Together, the results from these two experiments suggest that anticipation is not restricted to the simple cases of object motion typically examined in previous studies of representational momentum.

Nevertheless, several aspects of the current results deserve further comment. To begin with, the average memory shift of 40 ms corresponds to one video frame; however, our probe images were separated in four frame steps. We did this to ensure low responses in both tails of the distribution. In using relatively large spacing, we may have missed more subtle trends in the 100 ms surrounding the true/same stopping point, possibly underestimating the magnitude of the shift. Clearly, more fine-grained probe items could resolve this issue.

Next, we should return to the finding that for one of the four video clips there was no significant forward shift. Examination of Clip 4, the railway station sequence, shows that the amount of human activity in this clip is comparable to the other three clips, suggesting that ''lack of dynamics'' is not the answer. Of the four clips, however, the station scene does seem to provide the most natural options for static reference frames, both in terms of the train itself, running along the left of the shot and the various salient objects that form the right border of the shot. As we ran the present study in a blocked design, observers may have had ample opportunity to select one or more salient static items, making their same/different judgements relative to these points (see also Kerzel, 2002b). While we can only speculate on this issue, previous research has shown that such reference frames can significantly interact with (e.g., Hubbard & Ruppel, 1999) and even sometimes eliminate representational momentum shifts (e.g., Gray & Thornton, 2001). In Experiment 4, we try to reduce the likelihood of observers using such a strategy by running a mixed design in which different movie clips were interspersed with each other.

Finally, we come to the issue of how observers were actually performing their assigned task in this experiment. When we first conceptualized this experiment we had the rather naive notion that observers would be able to base their same/different judgements on an overall impression of change within a scene as a whole. However, both our subjective impression and those of our debriefed observers suggest instead that the most obvious strategy to adopt in this task is the selection of a subset of figures and/or background elements and to judge change in relation to the relative positions of these elements. In this respect, the task resembles the multielement experiments of Finke and Shyi (1988). Of course, it is not always a good idea to rely only on subjective report, and future studies could examine this issue more closely either by systematically reducing or restricting the available elements and/or by monitoring eye movements to determine the typical centres of interest in these displays.

This final point raises the issue of whether we are really tapping into the dynamics of the human figures at all, or whether observers are basing their decisions purely on the spatial location of people in the scene irrespective of what their targets' bodies are doing. We return to this issue in the next experiment.

EXPERIMENT 4

There were three main motivations for Experiment 4. First, we wanted to replicate the findings of Experiment 3 using the same task and stimuli and then to extend them by exploring whether a similar pattern would also be found with a new set of video stimuli. Second, we wanted to explore the possibility, mentioned above, that the blocked design used in Experiment 3 may have encouraged observers to focus on some aspect of the background content of the scene, using this as a reference frame. We speculated above that such a strategy could lead to more accurate localization as this had been observed in previous studies (e.g., Gray & Thornton, 2001). To reduce the possibility of observers using such a strategy, here we randomly intermix the scene presentation. That is, on a given trial in this experiment, observers were still not able to predict starting or stopping points of the moving objects (i.e. people), as in Experiment 3, but they are also unable to predict the static framework (i.e., the scene context) within which the moving objects appeared. Given such a brief presentation on each trial, it seems less likely that observers would make use of an unpredictable reference frame.

The third motivation for this experiment was to make an initial attempt at exploring the nature of the anticipation that might be taking place. While our goal was to study anticipation in the context of human action, our reflections on the likely strategy adopted by observers in Experiment 3 made it seem likely that they were focusing on the relative position of people in the scene, irrespective of the action being performed (for instance, the point of motion through the gait cycle).

As an initial attempt to address this issue, we included a condition in which the video clips were played backward. If the action-related content of the scene is irrelevant to the same/different decisions—as would be the case were observers simply using relative position of the figures—then performance should not differ between forward and reverse motion conditions. If on the other hand the meaning of the scene exerts an influence—as would be the case were observers at least somewhat sensitive to the actions—then we would predict some modulation of responses once the scene is shown backward. More specifically, we might predict that the magnitude of memory displacement would be greater in the forward-motion condition than in the reverse-motion one, when the clips are played out of real-world order. That is, if representational momentum is a reflection of anticipation, and if that anticipation is sensitive to predictability of actions, then the more familiar forward patterns should lead to greater anticipatory errors.

Methods

Participants. Sixteen observers from the Tübingen community were paid for participation in this study. All observers had normal or corrected to normal vision and were naive as to the purpose of the research until after the experimental session ended. None of the observers has taken part in Experiment 3.

Stimuli and equipment. The experimental set-up used here was identical to Experiment 3, with the exception that in addition to the four stimulus videos previously described (referred to below as Set A), we also introduced a second set, which are described below and will be referred to as stimulus Set B. As with Set A, an example movie of each scene can be viewed and/or downloaded from http://www.kyb.tuebingen.mpg.de/links/crowds.html.

- *Clip 5 (railway station, right view).* This clip is almost identical in content to Clip 4, except the camera is angled to the right to capture the opposite side of the station. The crowd density is slightly higher, with between 60 and 80 people visible along the length of the platform.
- *Clip 6 (shopping area alleyway).* This clip is taken in a busy alleyway between two shopping areas. Between 20 and 30 people move toward or away from the camera and within a range of 30–40 m. Walking speed is relatively slow (40 strides per minute). In one section of the clip a small number of people appear in relative close-up as they approach the position of the camera.
- *Clip 7 (town square, right view).* This clip is almost identical to Clip 2 above and shows crowds of between 90 and 150 people walking across a busy town square.
- *Clip 8 (department store exterior).* This clip was filmed from about 20 m distance from a department store exit. Crowd density is low, with only about 3–6 shoppers in view at any given time. A busy street can also been seen in the background of the shot.

Design. Each observer completed two conditions, forward and reverse motion, in separate blocks. The order of these blocks was counterbalanced with half of the observers completing the forward motion condition before the reverse motion and vice versa. Half of the observers saw Set A for the forward condition and Set B for the reverse condition (Group 1), while half saw Set B for the forward condition and Set A for the reverse conditions (Group 2). This gave a 2 (condition) × 2 (stimulus grouping) design. Within a block, the order of scene presentation was completely randomized on an observer-by-observer basis.

Task and Procedure. These were essentially identical to Experiment 3, except that only two training blocks were given: one to familiarize observers with the forward motion condition and one to familiarize them with reverse motion. As the content of the video clips was no longer blocked, a break was given after each 70 trials, to match the natural break point of Experiment 3.

Results

The solid line in Figure 5 shows the distribution of same responses for the forward motion condition, collapsed across sets of stimuli and observers. As in Experiment 3, this condition gave rise to a clear forward shift, with an average weighted mean of 34 ms, a figure that was reliably different from zero, $t(15) = 7.5$, $p < .001$. To examine whether this pattern was stable, separate one-way repeated measures ANOVAs were conducted for the two sets of stimuli. This revealed no significant main effect of video clip and, as indicated by the solid

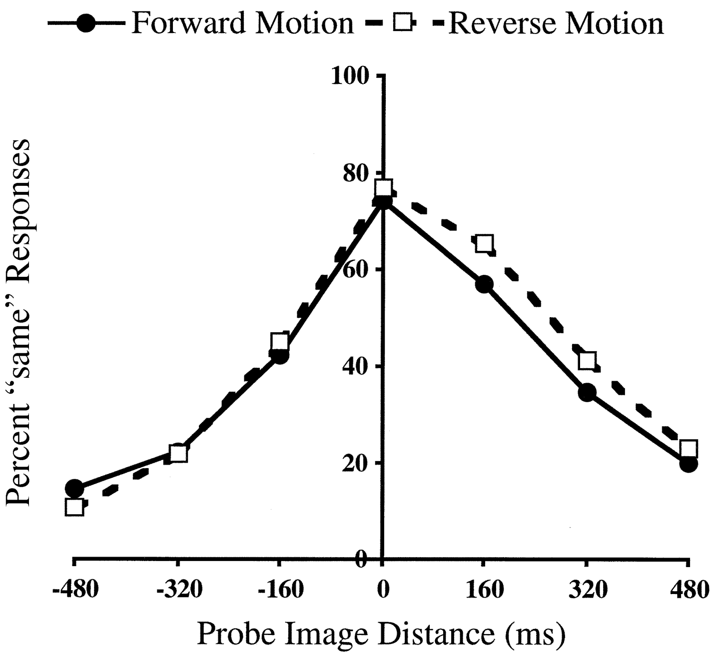

Figure 5. Percentage ''same'' responses, for forward motion and reverse motion video clips, as a function of probe distance in Experiment 4. Probe distance is measured in units of time along the video clip from the actual stopping point.

bars in Figure 6, there was little variation in the magnitude of the weighted means across any of the clips for the forward condition. In particular, the weighted mean for Clip 4, which had been much reduced in Experiment 3, is now of a similar magnitude to all the other clips.

Data from the reverse condition can be seen in the dashed line of Figure 5 and the open bars of Figure 6. There is a very similar pattern to that found with forward condition, although the magnitude of the shift is uniformly a little

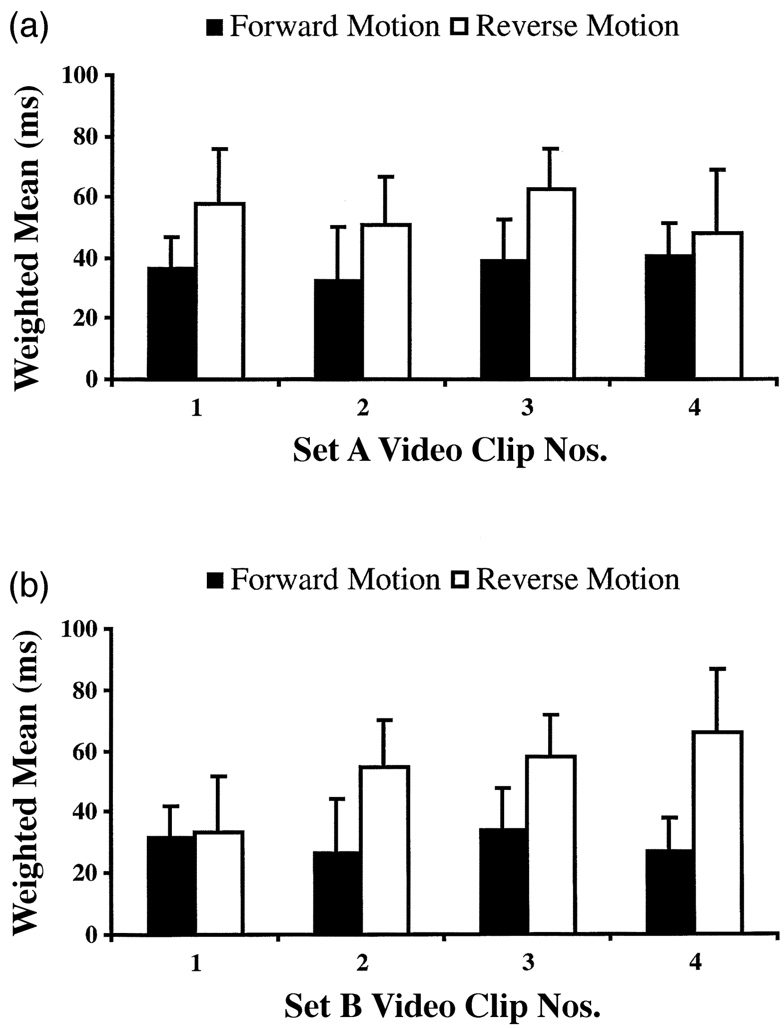

Figure 6. Estimated memory shifts for (a) Set A and (b) Set B in Experiment 4. Closed bars are for the condition with forward inducing motion; open bars for reverse motion.

larger. The overall weighted mean for the reverse condition, collapsed across stimuli sets and observers, was 54 ms, which again differed significantly from zero, $t(15) = 7.5$, $p < .001$. Again, separate repeated measures ANOVAs on the two sets revealed no significant differences between the individual clips.

To directly compare the pattern of results for the forward and reverse conditions, we conducted a 2 (condition) × 2 (stimulus grouping) ANOVA. The only reliable difference was the main effect of condition, $F(1, 14) = 5.521$, $MSE = 3173.798$, $p < .05$.

Discussion

The results of Experiment 4 essentially replicate those of Experiment 3, demonstrating the tendency for observers to endorse probe frames ahead rather than behind the true stopping point of the video sequences. Furthermore, our introduction of a mixed-trial design seems to have had the desired effect of reducing the opportunity for exploiting static reference frames as in this experiment all of the video sequences, including Clip 4—which had previously failed to show a bias—have similar magnitudes of forward shifts. Kerzel (2002b) had previously shown that randomized designs can lead to a reduction or even elimination of forward shifts compared to blocked designs. While the forward condition in Experiment 4 did give rise to a numerically smaller shift than in Experiment 3 (34 ms versus 40 ms), a direct comparison revealed that this difference was not significant, $t(20) < 1$, n.s. Possibly, if we had also randomized the direction of clip motion, rather than blocking it, a more pronounced reduction would have been observed.

The other main finding in Experiment 4 was a reliable difference between the forward- and reverse-motion inducing sequences. As the low-level motion cues, which would presumably be sufficient for action-independent localization, are identical regardless of motion direction, this finding suggests that some higher level aspect of scene processing (e.g., congruency of walking behaviour) is influencing responses. Unexpectedly, memory shifts were consistently larger for the reverse-motion video clips than for the forward-motion clips, contrary to our prediction.

One possible explanation for this pattern of results is that the allocation of attention is different for the two types of display. Hayes and Freyd (2002) have shown that reducing the attention allocated to simple displays, either by presenting multiple items or by introducing a secondary task, increases the amount of forward bias. It is conceivable that attentional processes are deployed more effectively to scenes that are familiar and meaningful compared to scenes consisting of the highly odd scenario of crowds of people moving backwards. The larger forward shift, then, would reflect less efficient processing of the reverse-motion displays. Clearly, this possibility needs to be investigated further, for example via eye movement analysis or dual task paradigms.

Regardless of the underlying mechanism, the difference in performance between forward- and reversed-motion video clips suggests that some aspect of individual motion of the walking figures is affecting performance. That is, the fact that people are articulating and translating backwards seems to make a difference. Had responses been based only on the overall layout and movement of abstract objects in the scene—with no reference to the individual actions—then no difference should have been observed.

GENERAL DISCUSSION

The four experiments reported in this paper have demonstrated that the forward shifts typically associated with representational momentum can be observed with complex, dynamic displays depicting human action. In Experiments 1 and 2, we used a simple virtual environment to simulate motion in depth along a scenic roadway. In Experiment 1, we found that memory for ego-position at the moment of a brief occlusion in the animation was shifted forward in the direction of motion. In Experiment 2, using a different task, we found that sensitivity to forward shifts in simulated position was consistently worse than sensitivity to backward shifts. Experiments 3 and 4 used short video clips of real-world crowd scenes to explore the accuracy of memory for observed action. In Experiment 3 we showed that memory was shifted forward in the direction of depicted actions. In Experiment 4, we replicated this finding and also found a consistent increase in the shift when the video clips were played backward, suggesting that anticipation can be modulated by scene content.

The most obvious contribution of the current set of findings is the generalization of representational momentum to displays that are considerably richer than those typically studied. The great promise of virtual reality and other technology, such as digital video, is that they allow us to more closely approximate real-world viewing conditions while still retaining sufficient experimental control. By going beyond displays with single translating or rotating objects, the current findings thus add empirical support to the theoretical suggestion that anticipation may be a very general feature of visual processing in the real-world (Freyd, 1987, 1993; Hubbard, 1995b).[2]

[2] Note that we would not want to suggest that anticipatory effects always dominate during visual processing. There have been a number of reported findings where memory shifts either do not appear at all or are biased against the direction of motion/change (e.g., Brehaut & Tipper, 1996; Hubbard, 1996; Kerzel, 2000; Thornton, 1997; Thornton & Vuong, 2002). It seems likely that such effects arise due to the dominance of other factors, such as the tendency to average or integrate across displays (Hubbard, 1996), the use of salient reference frames (Kerzel, 2000), or the desire to maintain object identity (Thornton, 1997; Thornton & Vuong, 2002). Understanding the relationship between such competing factors is clearly an important goal for vision science.

It is interesting to consider how the size of the memory shifts in the rich displays of the current experiments compare to the magnitude of forward shifts obtained in previous research using simpler displays. Although in past literature memory shifts have been reported in spatial units, these shifts can be converted to units of time, which provides a way of comparing the magnitude of memory shifts even when the dynamic and spatial contexts of the transformations are quite different. That is, given the speed of the stimulus in the inducing display we can calculate how long the stimulus would have travelled to produce a forward shift of a particular distance.

Accordingly, we calculated average memory shifts in units of time for two experiments previously reported in the literature that used similar methodology to that of the current experiments, more specifically, that used a retention interval of 240 or 250 ms and a probe method comparable to the current Experiments 1, 3, and 4.[3] For implied rotation in the picture plane Verfaillie and d'Ydewalle (1991, Exp. 1) reported a memory shift corresponding to a 34 ms shift; for implied translation in the picture plane Hayes and Freyd (2002, Exp. 2) reported shifts corresponding to a 68 ms shift under full-attention conditions and a 96 ms shift under conditions of divided attention. By comparison, in the current paper the 0.97 m memory shift associated with forward ego motion in Experiment 1 corresponds to a 52 ms shift; in Experiment 3 the forward shift was 40 ms, and in Experiment 4 the shifts were 34 ms for forward motion and 54 ms for backward motion.

In general, then, it seems that the memory shifts observed in complex displays are of the same order of magnitude as those observed in simple displays. Had the additional complexity completely eliminated representational momentum, then we would have sought explanations either in terms of reference frame effects (e.g., Gray & Thornton, 2001; Hubbard & Ruppel, 1999) or in terms of severe capacity limitations for representing complexity and dynamics. Had the size of the shift been greatly enhanced, then an explanation might have been framed in terms of the impact of divided attention or resources, in line with the work of Hayes and Freyd (2002). As it stands, our finding that memory shifts for complex, naturalistic displays appear to be neither greater nor less than shifts found for simple displays may indicate natural limits to the magnitude of representational momentum; clearly, such speculation can only be properly assessed once a wider variety of scenes have been tested.

Another contribution of the current work more directly relates to the theme of action. By showing that complex, action-related displays can give rise to representational momentum, we have opened the door for future studies where the observer has a less passive role. For example, similar to previous studies that

---

[3] When selecting experiments from the literature for this comparison, we limited our selection to papers in which the estimated memory shifts were reported numerically, as opposed to graphically.

have explored action in less complex scenarios (e.g., Jordan & Knoblich, in press; Jordan, Stork, Knuf, Kerzel, & Müsseler, 2001) we could compare memory for ego-position when the observer has active control of a simulated vehicle. Also, memory for observed actions that have recently been performed or are expected to be performed could be contrasted with the more passive viewing conditions used in Experiments 3 and 4. The direct manipulation of observer behaviour in this way may well lead to greater insights into the underlying relationship between vision and action in our spatial world.

In addition to the more theoretical implications of this work outlined above, we believe the current studies make several important methodological contributions. For example, it has recently been demonstrated that tracking eye movements could account for patterns of displacement in simple continuous motion displays (Kerzel, 2000). While we did not monitor eye movements in the current work, the general nature of the displays—multielements that smoothly move in depth as well as within the picture plane, often with brief display durations—makes it seem unlikely that nonanticipatory ''perceptual'' effects (Kerzel et al., 2001) are the main cause of the observed shifts. Similarly, the use of random stopping points and probe locations (Experiments 1–4) as well as the unpredictability of display content from trial to trial (Experiments 3–4) suggests that representational momentum is not as sensitive to precise task parameters as work with simple displays might suggest (Kerzel, 2002b).

Clearly, the work presented here is only a first step in exploring the anticipation of action in complex environments. Nevertheless, the current findings seem to indicate that our ability to visually predict the immediate future is not limited to situations involving simple, highly constrained displays.

REFERENCES

Actis Grosso, R., Stucchi, N., & Vacario, G. B. (1996). On the length of trajectories for moving dots. In S. C. Masin (Ed.), *Fechner Day 1996: Proceedings of the 12th annual meeting of the International Society for Psychophysics* (pp. 185–190). Padua, Italy: International Society of Psychophysics.

Bertamini, M. (2002). Representational momentum, internalized dynamics, and perceptual adaptation. *Visual Cognition*, *9*, 195–216.

Brehaut, J. C., & Tipper, S. P. (1996). Representational momentum and memory for luminance. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 480–501.

Daems, A., & Verfaillie, K. (1999). Viewpoint-dependent priming effects in the perception of human actions and body postures. *Visual Cognition*, *6*, 665–693

DeLucia, P. R. (2001). *Visual memory for moving scenes: Boundary extension or representational momentum?* Poster presented at the 42nd annual meeting of the Psychonomics Society, Orlando, FL.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General*, *129*, 481–507.

Faust, M. E. (1990). *Representational momentum: A dual process perspective.* Unpublished doctoral dissertation, University of Oregon, Eugene.

Finke, R. A., Freyd, J. J., & Shyi, G. C.-W. (1986). Implied velocity and acceleration induce transformations of visual memory. *Journal of Experimental Psychology: General*, *115*, 175–188.

Finke, R. A., & Shyi, G. C.-W. (1988). Mental extrapolation and representational momentum for complex implied motions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 112–120.

Freyd, J. J. (1983). The mental representation of movement when static stimuli are viewed. *Perception and Psychophysics*, *33*, 575–581.

Freyd, J. J. (1987) Dynamic mental representations. *Psychological Review*, *94*, 427–438.

Freyd, J. J. (1993). Five hunches about perceptual processes and dynamic representations. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 99–119). Cambridge, MA: MIT Press.

Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 126–132.

Freyd, J. J., & Johnson, J. Q. (1987). Probing the time course of representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 259–268.

Freyd, J. J., Pantzer, T. M., & Cheng, J. L. (1988). Representing statics as forces in equilibrium. *Journal of Experimental Psychology: General*, *117*, 395–407.

Fröhlich, F. W. (1923). Über die Messung der Empfindungszeit [Measuring the time of sensation]. *Zeitschrift für Sinnesphysiologie*, *54*, 58–78.

Futterweit, L. R., & Beilin, H. (1994). Recognition memory for movement in photographs: A developmental study. *Journal of Experimental Child Psychology*, *57*, 163–179.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gray, R., & Regan, D. (2000). Risky driving behavior: A consequence of motion adaptation for visually guided motor action. *Journal of Experimental Performance: Human Perception and Performance*, *26*, 1721–1732.

Gray, R., & Thornton, I. M. (2001). Exploring the link between time-to-collision and representational momentum. *Perception*, *30*, 1007–1022.

Hayes, A. (1997). *Representational momentum under conditions of divided attention.* Unpublished doctoral dissertation, University of Oregon, Eugene.

Hayes, A., Sacher, G., Thornton, I. M., Sereno, M. E., & Freyd, J. J. (1996). Representational momentum in depth using stereopsis. *Investigative Ophthalmology and Visual Science, 37*(Suppl. 3), S467.

Hayes, A. E., & Freyd, J. J. (2002). Representational momentum when attention is divided. *Visual Cognition*, *9*, 8–27.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding: A framework for perception and action. *Behavioral and Brain Sciences*, *24*(4), 869–937.

Hubbard, T. L. (1995a). Cognitive representation of motion: Evidence for representational friction and gravity analogues. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 241–254.

Hubbard, T. L. (1995b). Environmental invariants in the representation of motion: Implied dynamics and representational momentum, gravity, friction, and centripetal force. *Psychonomic Bulletin and Review*, *2*, 322–338.

Hubbard, T. L. (1996). Displacement in depth: Representational momentum and boundary extension. *Psychological Research/Psychologische Forschung*, *59*, 33–47.

Hubbard, T. L., & Bharucha, J. J. (1988). Judged displacement in apparent vertical and horizontal motion. *Perception and Psychophysics*, *44*, 211–221.

Hubbard, T. L., & Motes, M. A. (2002). Does representational momentum reflect a distortion of the length or the endpoint of a trajectory? *Cognition*, *82*, B89–B99

Hubbard, T. L., & Motes, M. A. (2003). Memory for the initial and final positions of a moving target: The Fröhlich effect, onset repulsion effect, and representational momentum. *Manuscript submitted for publication.*

Hubbard, T. L., & Ruppel, S. E. (1999). Representational momentum and landmark attraction effects. *Canadian Journal of Experimental Psychology*, *53*, 242–256.

Johansson, G. (1975). Visual motion perception. *Scientific American*, *232*(6), 76–88.

Jordan, J. S., & Knoblich, G. (in press). Spatial perception and control. *Psychonomic Bulletin and Review*.

Jordan, J. S., Stork, S., Knuf, L., Kerzel, D., & Müsseler, J. (2002). Action planning affects spatial localization. In W. Prinz & B. Hommel (Eds.), *Attention and performance XIX: Common mechanisms in perception and action* (pp. 158–176). Oxford, UK: Oxford University Press.

Kerzel, D. (2000). Eye movements and visible persistence explain the mislocalization of the final position of a moving target. *Vision Research*, *40*, 3703–3715.

Kerzel, D. (2002a). Different localization of motion onset with pointing and relative judgements. *Experimental Brain Research*, *145*(3), 340–350.

Kerzel, D. (2002b). A matter of design: No representational momentum without predictability. *Visual Cognition*, *9*, 66–80.

Kerzel, D. (2003). Mental extrapolation of target position is strongest with weak motion signals and motor responses. *Vision Research*, *43* (25), 2623–2635.

Kerzel, D., Jordan, J. S., & Müsseler, J. (2001). The role of perception in the mislocalization of the final position of a moving target. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 829–840.

Kirschfeld, K., & Kammer, T. (1999). The Fröhlich effect: A consequence of the interaction of visual focal attention and metacontrast. *Vision Research*, *39*, 3702–3709.

Knoblich, G., & Prinz, W. (2001). Recognition of self-generated actions from kinematic displays of drawing. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 456–465.

Kourtzi, Z., & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience*, *12*, 48–55.

Mather G., Verstraten, F., & Anstis, S. (Eds.). (1998). *The motion aftereffect: A modern perspective*. Cambridge, MA: MIT Press.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford, UK: Oxford University Press.

Munger, M. P., Solberg, J. L., Horrocks, K. K., & Preston, A. S. (1999). Representational momentum for rotations in depth: Effects of shading and axis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 157–171.

Müsseler, J., & Aschersleben, G. (1998). Localizing the first position of a moving stimulus: The Fröhlich effect and an attention-shifting explanation. *Perception and Psychophysics*, *60*, 683–695.

Müsseler, J., & Kerzel, D. (2003). Mislocalizations of the onset position of a moving target: Reconciling the Fröhlich and onset repulsion effects. *Manuscript submitted for publication.*

Nagai, M., Kazai, K., & Yagi, A. (2002). Larger forward memory displacement in the direction of gravity. *Visual Cognition*, *9*(1/2), 28–40.

Prinz, W., & Hommel, B. (Eds.). (2002). *Attention and performance XIX: Common mechanisms in perception and action*. Oxford, UK: Oxford University Press.

Thornton, I. M. (1997). *The perception of dynamic human faces.* Unpublished doctoral dissertation, University of Oregon, Eugene.

Thornton, I. M. (2002a). Further explorations of the onset repulsion effect. In M. Bauman, A. Keinath, & J. F. Krems, (Eds.), *Experimentelle Psychologie: Abstracts der 44. Tagung experimentall arbeitender Psychologen.* Regensberg, Germany: S. Roderer Verlag.

Thornton, I. M. (2002b). The onset repulsion effect. *Spatial Vision*, *15*, 219–243.

Thornton, I. M., DiGirolamo, G. J., Hayes, A., & Freyd, J. J. (1996). *Representational momentum under conditions of visual distraction*. Poster presented at the 37th annual meeting of the Psychonomic Society, Chicago, IL.

Thornton, I. M., & Hubbard, T. L. (Eds.). (2002). *Representational momentum: New findings, new directions.* Hove, UK: Psychology Press.

Thornton, I. M., & Vuong, Q. C. (2002, May). Representational momentum using complex, continuous motion. Poster presented at the second annual meeting of the Vision Sciences Society, Sarasota, FL.

Verfaillie, K., & Daems, A. (2002). Representing and anticipating human actions in vision. *Visual Cognition*, *9*, 217–232.

Verfaillie, K., de Troy, A., & van Rensbergen, J. (1994). Transsaccadic integration of biological motion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 649–670.

Verfaillie, K., & d'Ydewalle, G. (1991). Representational momentum and event course anticipation in the perception of implied periodical motions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 302–313.

Wetherill, G. B., & Levitt, H. (1965). Sequential estimation of points on a psychometric function. *British Journal of Mathematical and Statistical Psychology*, *18*, 1–10.