Psychonomic Bulletin & Review
2007, 14 (5), 989-993

Memory for moving and static images

W. J. Matthews
University of Warwick, Coventry, England

and

Clare Benjamin and Claire Osborne
University of Leicester, Leicester, England

Despite the substantial interest in memory for complex pictorial stimuli, there has been virtually no research comparing memory for static scenes with that for their moving counterparts. We report that both monochrome and color moving images are better remembered than static versions of the same stimuli at retention intervals up to one month. When participants studied a sequence of still images, recognition performance was the same as that for single static images. These results are discussed within a theoretical framework which draws upon previous studies of scene memory, face recognition, and representational momentum.

Work in the late 1960s and early 1970s established that recognition memory for pictures is exceptionally good (e.g., Shepard, 1967; Standing, 1973) and helped to initiate a fruitful program of research which has contributed to our understanding of the interplay between the perception, encoding and recognition of complex, meaningful visual configurations (see, for example, Henderson, 2005, for a survey of work on scene perception and memory).

Given that so much visual information involves movement, it is rather surprising that there has been very little investigation of long-term recognition memory for moving images. There has been substantial research into memory for the position of moving objects and the phenomenon of *representational momentum* (Freyd & Finke, 1984). Representational momentum (RM) is the systematic tendency to (mis)remember an event as extending beyond its actual endpoint (Thornton & Hubbard, 2002). Studies of RM have typically been concerned with the final positions of objects, and the time course of RM is limited to a few hundred milliseconds after stimulus offset (Freyd & Johnson, 1987). This work therefore differs from conventional studies of long-term recognition memory (although, as we shall argue below, investigation of RM is likely to provide important insights into the processes underlying long term recognition memory for moving images).

Similarly, although the effect of movement on recognition memory for faces has been quite extensively investigated (see O'Toole, Roark, & Abdi, 2002, for a review), we are aware of only one previous paper comparing long-term recognition memory for moving and static pictures, a preliminary investigation by Goldstein, Chance, Hoisington, and Buescher (1982). Goldstein et al. found that recognition memory was best when both the training and test stimuli were dynamic. This was contrary to their expectation that dynamic stimuli would be harder to remember "because static pictures contain less information than do the continuous visual stimuli" (p. 39). Goldstein et al.'s study represents a good first attempt at examining recognition memory for complex moving images, but suffered from the limited technology of the time. There was sometimes distortion in the stills taken from the moving clips, the duration of the clips shown in the test stage was only approximate, and during the recognition test all target stimuli occurred in the same order as in the original clip. The study also afforded no direct comparison of moving and static stimuli: the static images used during training were shown for 5 sec, in contrast to the 10 min of continuous film shown when the target stimuli were dynamic, and the 8-sec clips shown as targets and foils in the test stage. Since most studies of memory use a number of separate study items, it is desirable to employ the same approach with moving images by presenting participants with a set of discrete moving images.

With recent advances in the ease and accuracy with which computers can edit and display video material, the time is ripe to begin investigation of memory for moving images and scenes using the conventional paradigms of recognition memory research. The most obvious question to ask is whether the excellent recognition memory for static images represents an upper limit to human recognition performance, or whether memory for moving scenes is even better. There are conflicting predictions on this point. An ecological perspective (e.g., Gibson, 1979) would suggest that, since movement is the norm for visual stimuli, recognition memory for moving images ought to be better than for static images. Similarly, there is evidence that movement facilitates the learning of new faces (Lander & Bruce, 2003), and it may be expected that this advantage

W. J. Matthews, w.j.matthews@warwick.ac.uk

extends to more general stimuli. On the other hand, it may be that moving images contain more information and are therefore harder to encode (Goldstein et al., 1982).

We report two experiments in which memory for complex, moving scenes was compared with memory for static versions of the same images. Because any advantage/disadvantage for moving stimuli may be more obvious when the stimuli are degraded in some way (O'Toole et al., 2002) or after a certain retention interval, our first experiment compared recognition memory for moving and static black and white (i.e., relatively impoverished) and color images at one week and four week retention intervals.

EXPERIMENT 1

Method

Participants. Sixteen students from the University of Leicester took part.

Materials. A set of 800 digitally encoded video clips was assembled from a number of films. Relatively obscure, foreign film sources were chosen to reduce the probability of the participant having seen the film before; none of the participants in either experiment reported having seen any of the films from which the clips were taken. Varying numbers of clips were taken from each source, and care was taken to ensure that none of the clips were easily confused. All clips were 3 sec long and none contained cuts (sudden changes in scene or camera angle). All clips were played without sound. The clips depicted a wide range of scenes, including people, animals, vehicles, machinery, and natural views (e.g., the sea). The number of moving objects and the nature of the movement varied. A static image was obtained from each clip by taking a single frame from an arbitrarily chosen position in the clip.

Design and Procedure. Each participant studied 200 moving and 200 static images, half of each type being shown in color and half in black and white. Each stimulus stayed on screen for 3 sec, after which the participant rated its distinctiveness on a 6-point scale, ranging from "very distinct" to "very indistinct," by clicking the computer mouse on labeled boxes. This rating task was used to ensure engagement with the stimuli and to provide supplementary data, which are not reported here.

Participants returned one week after training and viewed 200 old and 200 new stimuli, of which 100 were moving, 100 static, with half of each type black and white and half color. Participants were told that they had seen some of these stimuli during the training session and that others were new, and they indicated whether or not they had seen each image in training. They returned again 4 weeks after training for a second test session, identical to the first. None of the stimuli shown in the second test had been seen in the first, and no participant saw both the moving and static versions of any clip. Similarly, none of the moving clips overlapped; that is, no clip had a start point drawn from within any other clip. (This structure was not explicitly stated in the task instructions, so it is possible that participants believed static images presented at test to have been embedded in moving clips during training, or that part of a moving image had been seen before but not all of it.)

In all stages, trials were organized into moving and static blocks, with monochrome and color stimuli randomly intermixed within blocks. The order of block presentation during training and test sessions was counterbalanced. Performance was measured by d'.

Results

The results are illustrated in Figure 1. A $2 \times 2 \times 2$ repeated measures ANOVA revealed that performance was better at the shorter retention interval [$F(1,15) = 26.7, p < .001, \eta_p^2 = .64$]; moving images were remembered better than static images [$F(1,15) = 9.02, p = .009, \eta_p^2 = .376$] and colored stimuli were remembered better than black and white ones [$F(1,15) = 26.9, p < .001, \eta_p^2 = .642$]. There were no significant interactions (all Fs < 1).

Discussion

The results demonstrate superior recognition memory for both black and white and color moving images relative to equivalent static images, an effect which is constant as retention interval increases from one week to one month. Consistent with previous work (Suzuki & Takahashi, 1997), color images were remembered better than black and white ones. The lack of interaction between chromaticity and movement suggests that the role of color in recognition memory for moving images is the same as for static images.

The superior recognition memory for moving images is consistent with ecological arguments (Gibson, 1979) and work on face learning (Lander & Bruce, 2003). As in Goldstein et al. (1982), our data do not support the suggestion that moving images are more difficult to remember because they contain more information.

Why are moving images better remembered than their static counterparts? One possibility is that the moving images provide more information about the objects contained in a scene, by showing them from several different angles. This idea is similar to the proposal that additional "perspective view information" forms the basis for superior recognition of moving faces (Pike, Kemp, Towell, & Phillips, 1997). A related possibility is that the total number of objects appearing in a moving image may be greater than in a static image (e.g., as objects move into frame). Alternatively, moving images may be better at capturing or holding the observer's attention; increased attention has also been suggested to underlie the superior learning of moving faces (Lander & Bruce, 2003).

To test these explanations, control stimuli were developed akin to the "multiple static" stimuli used by Lander and Bruce (2003) in the study of face recognition. For each moving clip, a series of six static images was obtained. Presenting these images in sequence gives a series of "snapshots" of the moving scene. If the advantage for moving images established in Experiment 1 arises because of additional perspective view information, or because the moving images contain more objects than the static ones, these multistatic stimuli should result in recognition performance which is superior to that for single static images. Similarly, if the moving images in Experiment 1 were better remembered because participants found them more interesting to study than their static counterparts, the multistatic stimuli would be predicted to be better remembered than the single static images, and perhaps as well as the moving ones, because the sudden appearance/disappearance and movement of objects from one image in the sequence to the next should attract attention and eye movements (e.g., Brockmole & Henderson, 2005). If, on the other hand, multiple static images are remembered no better than single static stimuli, alternative explanations for the superior recognition of moving images will need to be developed.

As well as introducing the multistatic stimuli, two minor procedural changes were introduced in Experi-

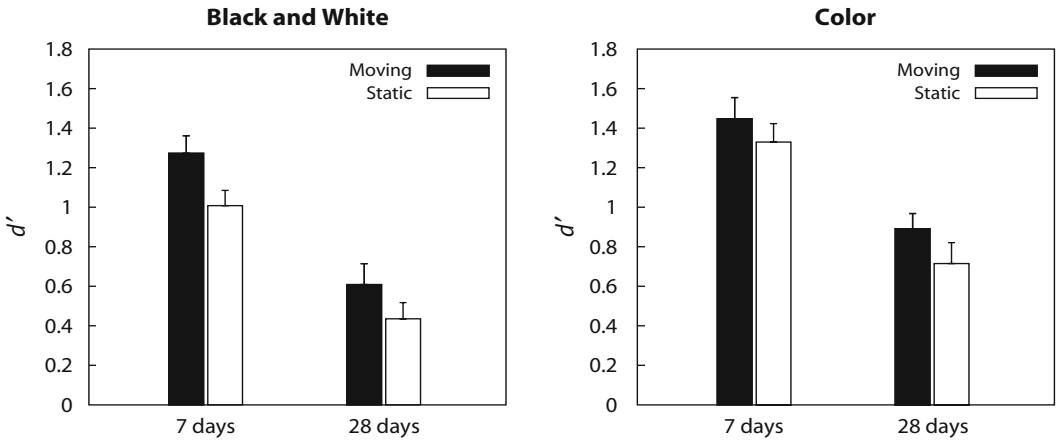

Figure 1. Results of Experiment 1, showing recognition memory for black and white (left panel) and color (right panel) moving and static images at 7-day and 28-day retention intervals. Error lines indicate plus one standard error of the mean.

ment 2 to test the generality of the findings from Experiment 1. First, rather than rating stimulus distinctiveness in the study phase, participants were instructed simply to study the stimuli in preparation for a subsequent memory test. Second, participants were tested at retention intervals of 1 day and 1 week.

EXPERIMENT 2

Method

Participants. Eighteen students from the University of Leicester took part, none of whom had participated in Experiment 1.

Materials. Three hundred of the 3-sec clips from Experiment 1 were used. For each, a multistatic stimulus was obtained by making stills 0 (the start of the clip), 0.5, 1.0, 1.5, 2.0, and 2.5 sec into the clip and showing each of these in order for 0.5 sec. Single static images were obtained by randomly selecting one of the still images making up each multistatic stimulus; each was shown for 3 sec.

Design and Procedure. During the study phase, each participant studied 50 moving clips, 50 multistatic, and 50 single static stimuli, blocked by type and with order of block presentation counterbalanced. They returned 24 h and 1 week later for test sessions in which they saw 25 old and 25 new stimuli of each type and indicated whether each had been seen during training; none of the stimuli shown in the first test session were used in the second. The order of moving, multistatic and single static blocks during the test stages was the same as in the study phase. As in Experiment 1, it was not explicitly stated that stimuli previously seen during training would only ever be shown in exactly the same form (static, multistatic, moving) at test, and that there was no overlap in the precise content of the stimuli.

Results

The results are illustrated in Figure 2. A 3 × 2 repeated measures ANOVA (using a Huynh–Feldt correction because of extreme violations of sphericity) indicated a significant effect of stimulus type [$F(1.51, 25.7) = 61.4$, $p < .001$, $\eta_p^2 = .783$] and better performance at the shorter retention interval [$F(1,17) = 29.6$, $p < .001$, $\eta_p^2 = .635$]. Stimulus type and retention interval did not interact [$F(1.2, 20.5) = 1.01$].

Paired t tests (corrected for multiple comparisons) showed that moving images were better recognized than both single static [$t(17) = 7.87$, $p < .001$, $d = 1.86$] and multistatic [$t(17) = 9.22$, $p < .001$, $d = 2.17$] stimuli. However, the multistatic and single static stimuli did not differ [$t(17) < 1$].

Discussion

The results of Experiment 2 support and extend those of Experiment 1; moving images were better recognized than still images taken from the moving clips, and the advantage for moving images was constant as retention interval increased. In the study phase of Experiment 1, participants rated the distinctiveness of each stimulus and were tested after retention intervals of 1 week and 4 weeks; in Experiment 2 participants studied the items in preparation for subsequent memory tests at 1 day and 1 week retention intervals. That the moving images were better remembered in Experiment 2 demonstrates that this effect is not due to the specific procedure employed in the first experiment, although it is noticeable that the recognition advantage

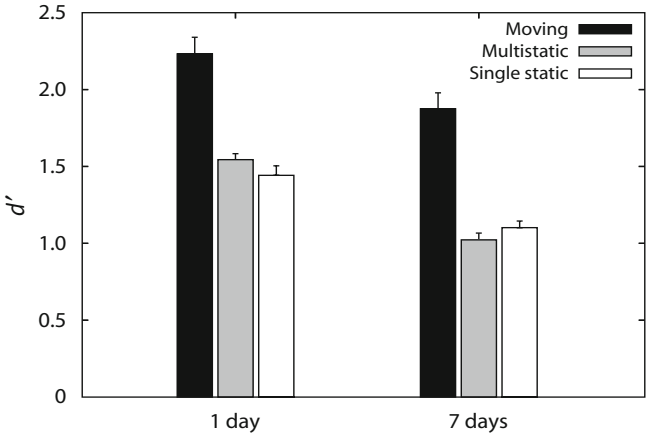

Figure 2. Results of Experiment 2, showing recognition memory for moving, multiple static (multistatic), and single static stimuli at 1-day and 7-day retention intervals. Error lines indicate plus one standard error of the mean.

of moving stimuli appears greater in Experiment 2 than in Experiment 1, which may be a consequence of these procedural differences.

More importantly, recognition memory for the multistatic stimuli was worse than for moving images and no different from memory for single static images. It therefore seems unlikely that the memory advantage for moving images is due to increased perspective view information, extra objects being present in the scenes, or greater attention being paid to the stimuli. Rather, our results suggest that fluid movement per se improves recognition memory; images with fluid movement have a memory advantage over stimuli containing much the same information but lacking fluid motion.

GENERAL DISCUSSION

Standing (1973, p. 210) suggested that, for vivid pictures, "memory capacity is almost limitless." The experiments reported here demonstrate that recognition memory for moving images is consistently superior to that for equivalent static images, and extend the preliminary findings of Goldstein et al. (1982).

A useful framework for considering the current results is the long-term object file theory proposed by Hollingworth and Henderson (2002) as a descriptive account of the visual processing of static scenes. According to this view, when eyes and attention are directed to a local object in a scene, visual processing establishes relatively "high-level" representations, which can include detailed information about visual form. These representations are indexed to the object's spatial position in the scene to form an object file. In Hollingworth and Henderson's model, the indexing of each abstracted visual and conceptual representation to a spatial position leads to storage in long-term memory as a *long-term memory object file*. Object files persist after attention is moved away from the object; as the viewer fixates on different positions in a scene, a relatively detailed representation of the overall scene is formed. Subsequent comparison of long-term memory representations with current, perceptual representations is dependent on the allocation of visual attention and gaze. That is, long-term memory retrieval is spatially mediated; changes to a target object are detected upon refixation. Furthermore, if a previously studied scene is encountered again sometime in the future, the scene map is retrieved and local object information can be retrieved by attending to the spatial position at which the object was originally encoded.

We suggest that it may be possible to modify the long-term memory object file model to incorporate moving images. Specifically, we propose that the map to which the positions of object representations are indexed is not only spatial, but spatiotemporal. In other words, when forming a memory of a dynamic scene or image, the viewer establishes a record not only of *where* a fixated object is, but also *when* it is there, and quite possibly where it is going. How might such spatiotemporal object files be formed? At this stage any suggestion is necessarily speculative, but the results of the current study, in combination with findings from studies of representational momentum (RM) and memory for moving faces, provide possible clues.

On the basis of studies of RM and boundary extension (BE, the phenomenon wherein participants overestimate the amount of a scene which was visible at encoding), DeLucia and Maldia (2006) have suggested that viewers use motion schemata to predict and extrapolate patterns of change so that eye movements and attention may be directed appropriately (e.g., to the likely current position of the object), and that it is the activation of these motion schemata which give rise to RM. The time course over which RM occurs is relatively short, typically peaking at approximately 200–300 msec. However, we suggest that since fixations are crucial to the formation and retrieval of object files (Hollingworth & Henderson, 2002), the activation of the short-term motion schemata underlying RM lead to patterns of directed attention and eye movements which support the formation of improved *long-term* memory codes in which abstracted object representations are indexed to spatial positions in the scene at particular times—that is, to the spatiotemporal scene map proposed above. Furthermore, activation of these motion schemata on subsequent presentation of the stimulus may guide attention and fixations in the same way as during encoding, thereby facilitating retrieval.

In Experiment 2, we found that memory for multistatic stimuli is the same as for single static images, which argues against the proposals that moving images are better remembered because they provide additional perspective view information or are better at capturing attention. This result may, however, be accommodated by the framework outlined above by assuming that the multistatic stimuli failed to activate motion schemata. Since motion schemata are hypothesized to guide eye movements and visual attention (DeLucia & Maldia, 2006), failure to activate these schemata will result in loss of information relevant to the construction of long-term spatiotemporally indexed object files. Presenting multiple static images may instead lead to the formation of separate object files/spatial maps for each image in the multistatic sequence, or perhaps to a representation of the scene in which each moving object is weakly associated with several different positions.

The suggestion that multistatic stimuli are less well remembered than moving stimuli because they fail to activate motion schemata initially seems contrary to evidence from studies of RM and BE using scene stimuli (DeLucia & Maldia, 2006; Munger, Owens, & Conway, 2005). These studies suggest that implied motion (IM) stimuli, which also consist of sequences of still images, can activate motion schemata. However, there is a potentially important difference between these IM stimuli and the multistatic stimuli used here: the former have blank interstimulus intervals (ISIs) between the successive static images. In Munger et al.'s study, for example, the IM stimuli consisted of three static images, each shown for 250 msec with a 250-msec ISI. In the multistatic condition of the current experiment, each static image was shown immediately after the preceding one, and the lack of a blank ISI may underlie the failure of the multistatic stimuli to activate motion schemata in Experiment 2.

Evidence consistent with this idea comes from Lander and Bruce (2003). Here, participants were asked to study faces moving nonrigidly (i.e., not simply turning the head from side to side) in either a continuous motion, multiple static or single static mode of presentation. As in our Experiment 2, the multiple static condition consisted of a sequence of images taken from the continuous clip and shown with no blank ISI between them. As in the present work, the continuously moving faces were remembered better than the multiple static and single static faces, which did not differ from one another. Therefore, just as the perception of apparent motion depends upon the precise spatiotemporal characteristics of the stimuli (e.g., Kolers, 1972), the lack of a suitable ISI may mean that the multistatic stimuli fail to activate appropriate motion schemata. Ecologically this makes some sense: Objects may change position while they are unattended or obscured, but do not instantaneously disappear from one position and appear in another. In addition, there is further empirical evidence consistent with the idea that the specific spatiotemporal structure of the stimuli influences activation of motion schemata and thence encoding of the moving image in long-term memory. Lander, Christie, and Bruce (1999) compared memory for faces moving normally, with slowed motion and with disrupted, jerky motion. The normal movement led to best recognition performance; memory in the slowed and jerky conditions was inferior. It seems plausible that, like the multistatic stimuli used in our Experiment 2 and by Lander and Bruce, the jerky- and slow-motion stimuli did not activate motion schemata, resulting in poorer recognition memory.

The empirical results and theoretical framework outlined here suggest several lines for future enquiry. The idea that multiple static stimuli and jerky film sequences fail to trigger the motion schemata activated in studies of representational momentum may readily be tested by examining RM with these stimuli. Similarly, the continuously moving images used in the experiments reported here (and in studies of RM and memory for faces) actually consist of a sequence of still images shown rapidly enough to create the appearance of fluid movement. Employing additional multiple static conditions with varying numbers of still images and different frame rates will therefore help to pinpoint the conditions for motion schema activation. Exploring the validity of our more general suggestion that moving scenes are encoded as long-term spatiotemporal object files will require eye-tracking studies akin to those using static scenes. For example, it will be important to ask how fixation on a moving object influences its long-term representation in memory for the scene (cf. Hollingworth & Henderson, 2002), and how important the semantic consistency of movement is for memory formation (Henderson, Weeks, & Hollingworth, 1999). We hope that the present study will serve as a starting point and impetus for future research in this area.

AUTHOR NOTE

We thank Margaret Munger and Carrick Williams for their helpful comments on an earlier version of this article. Correspondence concerning this article should be addressed to W. J. Matthews, Department of Psychology, University of Warwick, Coventry CV4 7AL, England (e-mail: w.j.matthews@warwick.ac.uk).

REFERENCES

Brockmole, J. R., & Henderson, J. M. (2005). Prioritization of new objects in real-world scenes: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception & Performance*, **31**, 857-868.

DeLucia, P. R., & Maldia, M. M. (2006). Visual memory for moving scenes. *Quarterly Journal of Experimental Psychology*, **59**, 340-360.

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

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

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

Goldstein, A. G., Chance, J. E., Hoisington, M., & Buescher, K. (1982). Recognition memory for pictures: Dynamic vs. static stimuli. *Bulletin of the Psychonomic Society*, **20**, 37-40.

Henderson, J. M. (Ed.) (2005). *Real-world scene perception.* Hove, U.K.: Psychology Press.

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 210-228.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception & Performance*, **28**, 113-136.

Kolers, P. A. (1972). *Aspects of motion perception*. New York: Pergamon.

Lander, K., & Bruce, V. (2003). The role of motion in learning new faces. *Visual Cognition*, **10**, 897-912.

Lander, K., Christie, F., & Bruce, V. (1999). The role of movement in the recognition of famous faces. *Memory & Cognition*, **27**, 974-985.

Munger, M. P., Owens, R. T., & Conway, J. E. (2005). Are boundary extension and representational momentum related? *Visual Cognition*, **12**, 1041-1056.

O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences*, **6**, 261-266.

Pike, G. E., Kemp, R. I., Towell, N. A., & Phillips, K. C. (1997). Recognizing moving faces: The relative contribution of motion and perspective view information. *Visual Cognition*, **4**, 409-437.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning & Verbal Behavior*, **6**, 156-163.

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, **25**, 207-222.

Suzuki, K., & Takahashi, R. (1997). Effectiveness of color in picture recognition memory. *Japanese Psychological Research*, **39**, 25-32.

Thornton, I. M., & Hubbard, T. L. (2002). Representational momentum: New findings, new directions. *Visual Cognition*, **9**, 1-7.

(Manuscript received July 28, 2006;
revision accepted for publication November 7, 2006.)