Perception and memory across viewpoint changes in moving images

Yoriko Hirose
School of Psychology, University of Dundee, Dundee, Scotland, UK

Alan Kennedy
School of Psychology, University of Dundee, Dundee, Scotland, UK, & Laboratoire de Psychologie et Neurosciences Cognitives, Boulogne-Billancourt, France

Benjamin W. Tatler
School of Psychology, University of Dundee, Dundee, Scotland, UK

Current understanding of scene perception derives largely from experiments using static scenes and psychological understanding of how moving images are processed is under-developed. We examined eye movement patterns and recognition memory performance as observers looked at short movies involving a change in viewpoint (a cut). At the time of the cut, four types of object property (color, position, identity and shape) were manipulated. Results show differential sensitivity to object property changes, reflected in both eye movement behavior after the cut and memory performance when object properties are remembered after viewing. When object properties change across a cut, memory is generally biased towards information present after the cut, except for position information which showed no bias. Our findings suggest that spatial information is represented differently to other forms of object information when viewing movies that include changes in viewpoint.

Keywords: eye movements, visual memory, scene perception, motion, movie, representation

Introduction

This paper is concerned with how object information is represented in moving images that contain viewpoint changes, and how such changes influence subsequent eye movement behavior and recognition memory for objects. In movies we are often faced with sudden changes in viewpoint (editorial cuts). These abrupt cuts in films often go unnoticed by the observer (Smith & Henderson, 2008) and may even aid the narrative development of the film. Editorial cuts occur every 2.7–5.4 seconds in a typical Hollywood film (Bordwell & Thompson, 2001). The perceptual consequences of cuts are of interest both in terms of how filmed sequences are understood and how we extract and integrate meaningful information across changes in the point of view presented to the observer.

The perceptual and memorial consequences of interruptions during viewing have been the focus of much recent research and these may provide insights into how we encode information across cuts in films. Early change detection studies (Grimes, 1996; see Simons & Rensink, 2005, for a review) suggested that the ability to detect change across saccades, blinks and artificial interruptions was poor. The initial response was the claim that the internal representation of the relevant visual information must be, at best, very sparse (Rensink, O'Regan, & Clark, 1997; Simons & Levin, 1997) or that no such second-order representation existed (e.g. O'Regan, 1992; O'Regan & Noë, 2001). More recent work however has cast doubt on this position from two directions (see Carmi & Itti, 2006, for further discussion). First, there is a methodological weakness in the claim that lack of conscious change detection necessarily precludes the existence of some form of internal representation (Simons & Rensink, 2005). Overt change blindness may co-occur with evidence of implicit knowledge of change (Fernandez-Duque & Thornton, 2000; Silverman & Mack, 2001). Second, a substantial body of empirical evidence suggests that relatively detailed scene information can be preserved across fixations from static scenes (Hollingworth & Henderson, 2002; Irwin & Zelinsky, 2002; Melcher, 2006; Tatler, Gilchrist, & Rusted, 2003).

However, a significant limitation of most work on memory for object properties has been that the results were derived from experiments involving the viewing of static scenes, often with explicit memory tasks using simplified stimuli. Memory processes under these conditions may not match those under more realistic settings (Tatler, Gilchrist, & Land, 2005; Tatler & Melcher, 2007).

Moreover, the real visual environment is anything but static. Any representation constructed to serve natural behavior must be flexible enough to withstand both changes in the environment and the changes in viewpoint that result from observer movement. Dynamic scene stimuli offer an interesting compromise as experimental stimuli that bridge the gap between static scene viewing and natural behavior. Moving images often contain changes in represented viewpoint brought about by changes in camera position, which are similar to those resulting from observer movement in the real world, though more abrupt than would occur through ego-motion. Significant aspects of eye movement control in the real world will not be captured in the viewing of moving images, but film manipulations do have the advantage of allowing for the necessary degree of control demanded by psychological experimentation. Understanding the way observers use dynamic images to construct a coherent internal representation is an important step towards improving our knowledge of the processes that operate in the more complex real-world setting.

Until recently, work on the perception and memory for object properties in moving images was restricted to studies using dynamic stimuli to examine change detection. For example, Levin and Simons (1997) found that changes made to object color, presence, position and identity during cuts were rarely detected. Although a variety of changes were made during a cut, this study did not systematically explore how different object properties were remembered across changes in viewpoint, or how particular changes influenced behavior and memory. These issues were addressed by Wallis and Bülthoff (2000), using movies shot in virtual environments simulating observer motion. The results suggest that the detection of orientation, position and, to a lesser degree, color changes was worse during (simulated) observer motion than in static viewing of the same scene. In contrast, the mode of presentation had little effect on the detection of presence per se. Angelone, Levin, and Simons (2003) also found that color changes across a cut in short movies were often not detected, even though observers recognized the pre-change stimulus color at an above chance level. On the basis of these studies, however, it is not possible to answer an important prior question: whether visual information is integrated across changes in viewpoint when object properties are unchanged across a transition in moving images. When movies contain no editorial cuts memory performance is actually better in dynamic than in static scenes. Matthews, Benjamin, and Osborne (2007) compared recognition memory performance for movies with no editorial cuts, single static images and sequences of still images taken from the same movies, at retention intervals of up to a week. Moving images were remembered better than the two static versions, which were equally well remembered. Thus, movement itself may be beneficial in representing scene

information. A follow-up study by Buratto, Matthews, and Lamberts (2009) found that recognition accuracy was highest when study and test format were the same, showing a congruency effect. However these studies only tested recognition memory and no detailed memory for scene content was tested. It is yet to be established how different object properties are represented in dynamic stimuli that contain viewpoint changes.

Recently, attention has shifted from the analysis of detection performance to the interaction between changes in viewpoint and perceptual memory. Carmi and Itti (2006) examined bottom-up and top-down influences on attentional selection during dynamic scene viewing using video clips containing abrupt transitions (jump cuts). When a completely novel scene was introduced there was a greater correlation between low-level image properties and fixation selection. Since this correlation diminished over time, one possible interpretation is that the influence of top-down effects becomes stronger after scene gist and layout have been recognized. The effect of viewpoint changes on recognition memory in dynamic scenes has also been examined. Garsoffky, Schwan, and Hesse (2002) used video clips of football episodes taken from one or two viewpoints and tested recognition memory using stills from the video varying in viewpoint. Recognition accuracy was higher when tested using the same viewpoint, suggesting that representations may, to some degree, be viewpoint-dependent. However, the results of Garsoffky, Huff, and Schwan (2007), using computer-animated basketball scenes shown from one or two viewpoints, found that recognition accuracy did not vary as a function of viewpoint. It is unclear whether the use of simplified animated stimuli contributed to this difference and more work is needed to examine the role of viewpoint deviation on recognition of moving images. Furthermore, no studies to date have attempted to probe for memory for different kinds of visual information during viewing of dynamic scenes that contain viewpoint changes. The present paper attempts to remedy this by examining memory for particular object properties as observers watched short films (henceforth termed movies) containing cuts. Specifically, we examined memory for four types of object property in movies: color, position, identity (type) and shape (token) of the target object. Our aim was to systematically examine the representational processes for the four types of object property when viewing movies containing viewpoint changes.

Prior studies using static scene stimuli have shown that memory for the four object properties listed above survive the saccade (Henderson & Hollingworth, 2003; Hollingworth & Henderson, 2002; Irwin & Zelinsky, 2002; Melcher, 2006; Melcher & Kowler, 2001; Tatler et al., 2003, 2005). In natural behavior detailed object information is also represented during action, although only task-specific information required at the present moment is extracted (Ballard et al., 1992; Ballard, Hayhoe, & Pelz, 1995; Droll, Hayhoe,

Triesch, & Sullivan, 2005; Triesch, Ballard, Hayhoe, & Sullivan, 2003). Thus the nature of action goals seems to be an important factor in determining what information is represented during natural behavior. How is object information encoded and integrated across viewpoint changes in moving images, and are all types of object property represented equally well? Movie sequences appear to be processed without effort, yet the fact that people are rather poor at detecting changes that happen during an editorial cut (e.g. Levin & Simons, 1997) suggests that extraction and integration of object information across viewpoint changes may be difficult to achieve in moving images. In particular, representing position information in movies may be more difficult than processing other types of information such as color or identity. In static scene viewing object position is coded in relation to the locations of other objects in a larger spatial representation, which provides a contextual frame of reference (Hollingworth, 2007). In dynamic scenes, spatial information may be encoded in much the same way, coding object positions with respect to a larger external frame of reference (such as the screen in which the movie is shown) or with respect to other objects in the scene. However, in such dynamic scenes, coding object position in a representation of the scene requires constant updating of information with respect to changing external frame of reference. This need for constant updating is in contrast to the situation for coding color or identity because these object properties are independent of background or external factors. Moreover, a number of studies have suggested that position information may be represented in a qualitatively different manner from other object information. Position information may be extracted before other sources of information in order to construct an overall spatial layout of the scene (Aginsky & Tarr, 2000; Rensink, 2000; Tatler et al., 2003). Similarly, how position information is accumulated over fixations has been found to be distinct from other object properties. In static scenes memory for position appears to accumulate over successive fixations (Melcher, 2006; Tatler et al., 2005), whereas the accumulation of color and identity information appears less consistent, with some studies observing no accumulation (Tatler et al., 2005) and others showing an opposite pattern (Hollingworth & Henderson, 2002; Melcher, 2006). In moving images we may therefore expect to find differences in the sub-structure of visual representations for different object properties, in particular between those properties that can be coded independent of external factors (e.g. color, identity) and property such as position that requires some form of external frame of reference.
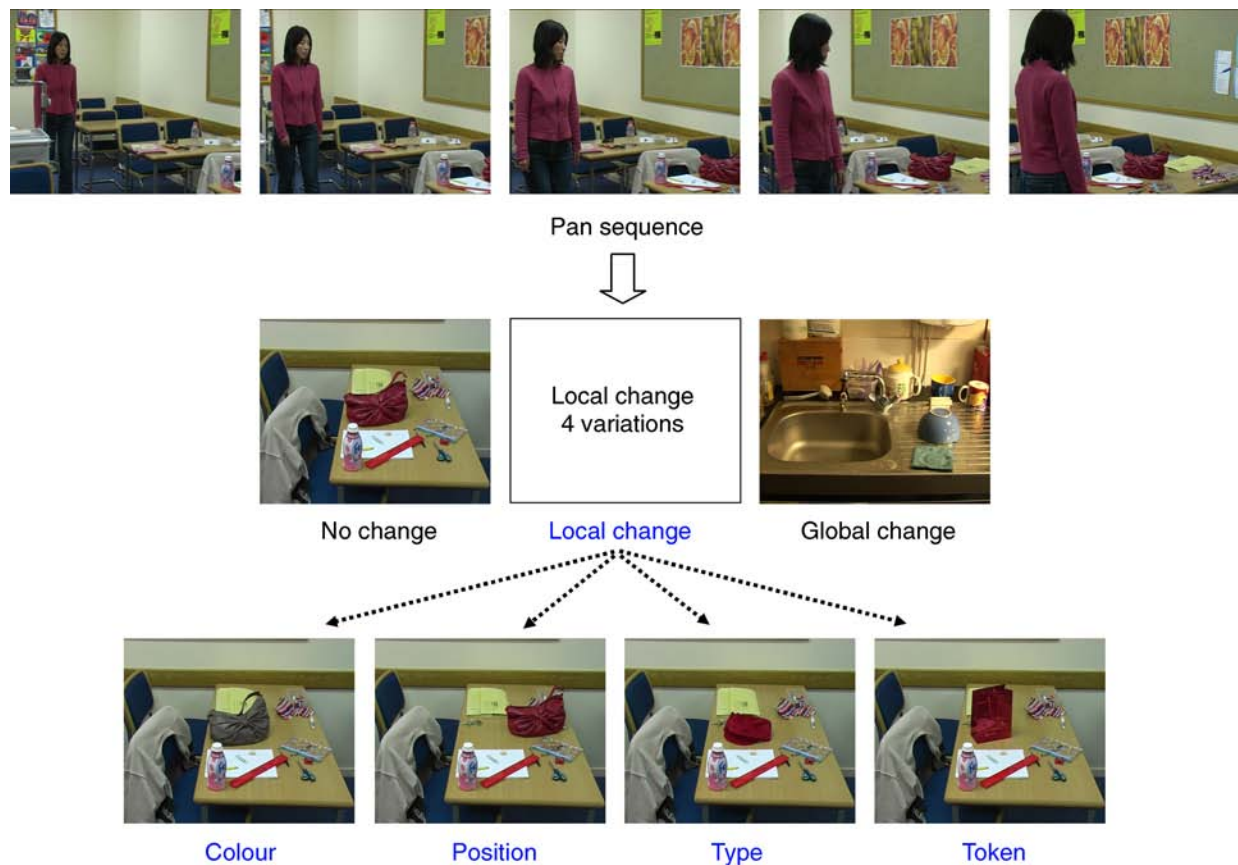

Figure 1. An illustration of a movie. The example shows the scenes shot in the "lecture room", with the "bag" as the target object.

The movies used in the current study all had the same structure (see Figure 1 and Method). The camera maintained an image of an actor walking through a scene. This was a conventional "pan shot" in which the actor's image was roughly stable in the center of the screen and background objects were in relative motion. In all cases the pan shot ended with the actor coming to a halt, at which point there was a cut to another viewpoint. This defined a critical transition and was followed by an image filmed from a static viewpoint. In some conditions, the post-cut image represented a plausible view of the imaged scene "through the actor's eyes". In other conditions the cut was to a completely different scene. The principal experimental manipulations related to systematic changes in properties of a single object visible during the pan-shot and, in some cases, visible after the cut, albeit from a changed viewpoint. When a change occurred, other aspects of the scene were unchanged across the cut. Observers' eye movements were recorded continuously. The experimental work was designed to comment on two broad research questions. First, the degree to which changes in inspection behavior around the time of a cut reveal differential sensitivity to changes in particular object properties (color, position, type and token). This question was addressed by an examination of eye movements around a cut as a function of whether object properties changed or remained unchanged. Second, the degree to which the characteristics of object information extraction and retention inferred from static scene viewing and natural behavior can be extended to dynamic scene perception. This was addressed by analyzing the patterns of recognition memory for the four types of object property. The specific experimental hypothesis examined in the context of these research questions relates to the sub-structure of visual representations. Following previous findings from static scene viewing (e.g. Hollingworth & Henderson, 2002; Melcher, 2006; Tatler et al., 2005), we predict that different types of object properties are also encoded and retained to different degrees in moving images. In particular, we expect that position information, requiring external frame of reference, will be represented in a way distinct from color or two types of identity information (type and token) that require no contextual frame of reference.

Method

Participants

Sixty participants (mean age 23.6 years, *SD* = 7.3), randomly divided into four equal groups, participated in the study. All had normal or corrected-to-normal vision and were naïve to the purposes of the study and were paid five pounds for taking part. Written informed consent was obtained from all participants in accordance with the University of Dundee Research Ethics Committee Regulations on research involving human participants.

Stimuli

Stimuli consisted of fifteen movies (no audio), recorded in several locations at the University of Dundee (e.g. bookshop, café, lecture room). Movies were recorded using a Canon XM2 mini-DV camcorder and subsequently edited using Adobe Premiere Pro 2.0. Each movie lasted 10 seconds (25 frames per second, 720×576 pixels per frame). Each started with a panned sequence (5 seconds) in which an actor moved through an environment containing several objects and came to rest near a target object (see Figure 1). On average, the target object was visible for around 3 seconds during the pan (76 frames; 125 frames maximum). Unique target objects were allocated to each scene (e.g. folder, jug, bag). The pan was followed by a five-second sequence in which the camera was maintained in a static viewing position. Three types of post-cut scene were created. In a baseline control condition (termed "no change") there was a cut of approximately 90°, plausibly reflecting the direction of the actor's gaze immediately before the cut. No object properties were changed. A second control condition (termed "global change") involved a cut to a completely different scene. These were recorded in several locations at the University of St. Andrews (e.g. foyer, reception, workshop), and generally contained fewer objects than the scenes used in the no change condition. In the experimental condition (termed "local change") a single change was applied to a target object in the post-cut scene. In different experimental conditions, one of four types of change was made to the target object: a change in color, position, identity (a change to object type) or shape (a change to the particular object token). All changes were physical changes, made using real objects in the scene rather than post-production editing.

Figure 1 shows the four types of change in a "lecture room" movie with "bag" as the target object. In the color change condition, the color of the target object was changed by replacing a red bag with an otherwise identical gray bag. In the position change condition, a change was applied to the location of the target object relative to its locally defined context, by physically moving the target to a different location relative to other objects (e.g. it was displaced to another location on the table). In the type change condition the target object was replaced by an object of a different kind but the same color and of a similar size as the target (e.g. a red bag was replaced by a red hat). Finally, in the token change condition the target object was replaced by a different exemplar of the same object class in the same color and of similar size as the target (e.g. a red gift bag). See Movie 1 for examples of the movies used in the study.

Procedure and design

Prior to the experiment, participants were informed that they were going to view several video clips and answer questions regarding the content of the movies. They were

Movie 1. Examples of the movies used in the study.

informed that in some cases they might notice an error in continuity between the pan and the post-cut scenes ("Some videos contain editing errors and you may notice unusual scenes"). Participants were first given a practice trial to familiarize themselves with the procedure, then viewed the fifteen video clips in random order. The four types of change were tested in four separate experiments (N = 15 in each). Participants in each group saw five scenes in each post-pan condition (no change, global change, local change). Fifteen scenes were counter-balanced using a Latin-square design with three versions of the test sets.

After viewing each movie, participants answered a question regarding the presence of a continuity error ("Did you notice anything unusual in this video clip?") and four four-alternative forced choice (4AFC) questions regarding the visual properties of the target object. Specifically, these questions mirrored the four types of change (color, position, type, token) manipulated in the study. The first 4AFC question tested participants' knowledge of the target object regarding its presence in the scene (type information) using category names in words. The second question tested knowledge about the color of the target object using color patches. In the third question, participants were asked to indicate the position of the target object, with reference to a simple line-drawing of the post-cut scene. In the final question, participants were asked to identify the target object (token information) from an array of grayscale images of real objects. The questions presented to participants after viewing the "lecture room" video clip are shown in Figure 2.

In all four 4AFC questions, one answer always matched the visual property of the target object in the pan and another matched the property of the target used in the post-cut scene in the four corresponding change conditions (color, position, type, token). The two remaining answers were chosen to represent plausible fillers appropriate for

each scene. For example, after viewing the "lecture room" video clip (Figure 1), participants answered a question related to the type of the target (Figure 2, top left). Here, two answers ("bag" and "hat") were the target object shown during the pan and the changed target in the type change condition, respectively. The other two answers ("newspaper", "scarf") were objects that were appropriate for the scene, in this case a lecture room. The same rule was applied to the other three 4AFCs when selecting the fillers.

It should be noted that the four types of change were, of necessity, tested in four separate experiments. Thus, observers in any given experiment experienced only one type of local change. That is, in each experiment only one of the four 4AFC questions was relevant to the local change being examined. This means that in the remaining three 4AFC questions there was always one correct answer and three fillers novel to the observers (although one of the fillers, in fact, matched the post-cut target property in other experiments). Depending on the type of the post-cut scene, the ratio of correct answers and fillers differed in the question related to the local change. After a local change, the question contained two possible correct answers (based on information before and after the cut) and two fillers. Following the two control conditions (no change and global change) there was only one correct answer. Complete descriptions of all the questions and movies used are shown in Appendices A and B.

Stimuli were displayed on a ViewSonic P225f 22″ pure flat CRT monitor running at a refresh rate of 85 Hz. The monitor was positioned at a viewing distance of 60 cm and the movies therefore subtended 28.1° × 22.5° of the observer's visual field. Binocular eye movements were recorded using an SR Research Ltd. EyeLink II eye-tracker (500-Hz sampling rate, ±0.5° accuracy) with compensation for head movements. A 9-point target array was used to calibrate eye position and a second 9-point array then used to validate the calibration and compute the mean spatial accuracy of the eye-tracker calibration. If the second 9-point array revealed a spatial accuracy worse than ±0.5° the eye-tracker was re-calibrated. Eye position data were collected for the eye that produced the better spatial accuracy as determined by the calibration. Before each trial, a drift correction was performed to maintain spatial accuracy. Saccades and fixations were defined using the saccade detection algorithm supplied by SR Research. Saccades were identified by deflections in eye position in excess of 0.1°, with a minimum velocity of 30°sec-1 and a minimum acceleration of 8000°sec-2, maintained for at least 4 ms.

Methodological precautions

If we wish to make relative statements about how well particular object properties are remembered, or how sensitive participants are to different types of object
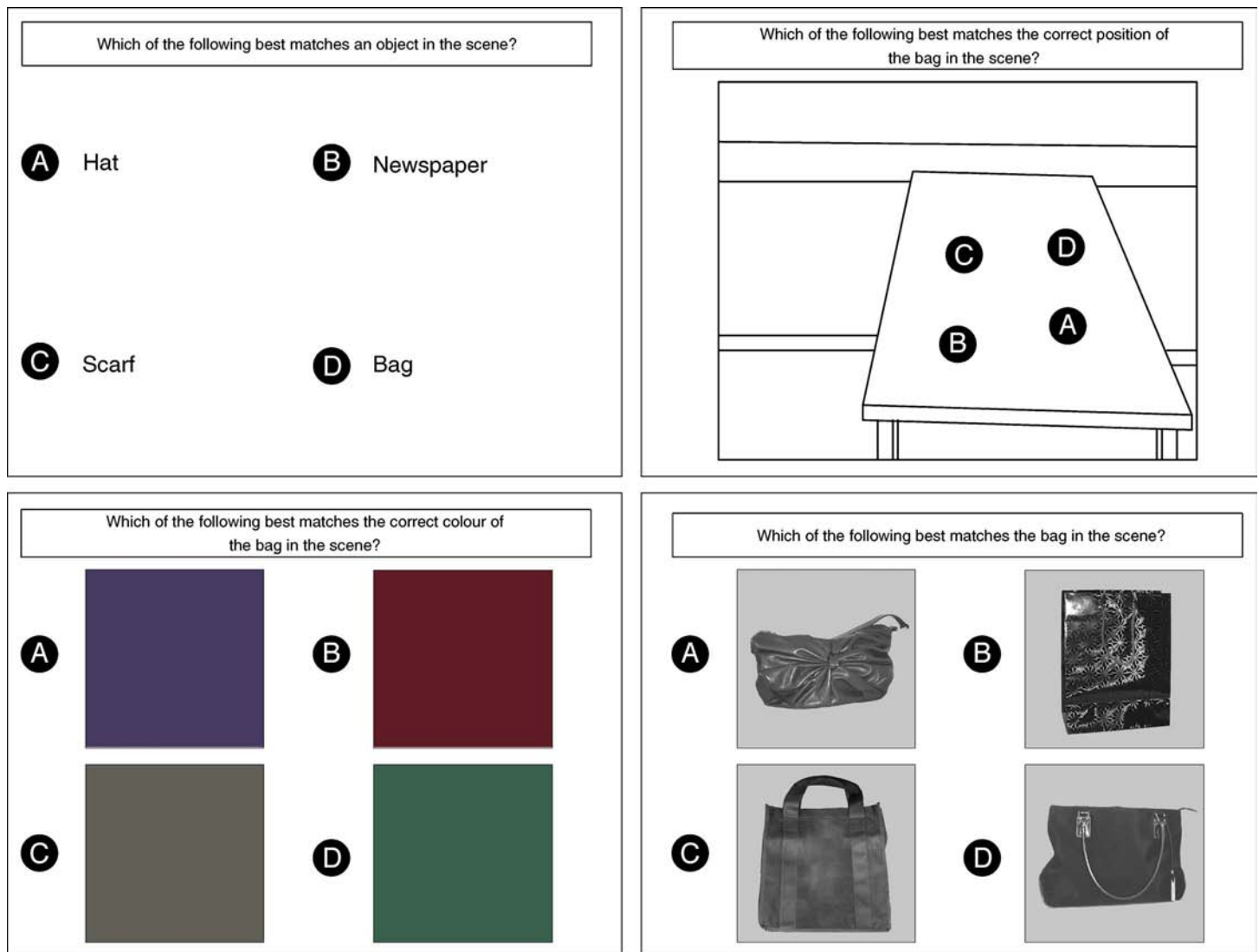
Figure 2. An illustration of the four-alternative forced choice questions given to participants after viewing the "lecture room" movie. Participants also answered a question regarding the presence of a continuity error ("Did you notice anything unusual in this video clip?") before answering the four-alternative forced choice questions.

change, then the questions must be carefully designed and balanced. In particular, every effort must be made to reduce and control the degree to which some types of change might be inherently and perceptually more detectable than others. For example, one should seek to ensure that the degree of change between two chosen colors is about equal to the degree of change between two chosen shapes. However, achieving this balance in practice is not straightforward.

One objective measure of the perceptual discriminability of objects is to use the salience model developed by Laurent Itti and colleagues (e.g. Itti & Koch, 2000). This model uses a competitive combination of local color-intensity and orientation-contrast in images in order to compute an overall conspicuity map of the image. This objective measure of visual salience can then be used to consider whether the conspicuity of the target objects in

the scene differed depending upon the type of object property change made during the editorial cut. We computed salience maps for each of the four types of object property change experienced after the editorial cut, and also for the scene as viewed at the end of the panned sequence. Each salience map was calculated using the latest version of the Salience Toolbox for Matlab, available at http://www.saliencytoolbox.net (downloaded on 18th January 2010). We used the default parameter settings for computing salience maps (for details of the algorithm see Walther & Koch, 2006). We used the computed salience maps to consider the perceptual conspicuity of the (changed) target items after undergoing each type of change by defining the object's location with a circle surrounding the object as closely as possible.

The results show the target object rarely coincided with any of the first five most salient locations in the scenes

(with one of the first five most salient locations overlapping with the object in an average of 18.3% of cases across the four types of object property change). A one-way ANOVA revealed no significant effect of change type upon the frequency with which one of the first five most salient locations overlapped with the object, $F(3, 42) = 0.75$, $p > 0.1$. We also used the graded output of the salience map (normalized by dividing by the maximum value on the salience map in order to make the salience maps of the different scenes comparable) and considering the maximum value of salience within the boundary defined around the target object. Including the final view of the panned sequence in our one-way ANOVA, we found no main effect of condition (end of pan, color change, position change, type change, token change) on salience values within the object boundary, $F(2.53, 35.37) = 1.37$, $p > 0.1$ (the degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity, $\varepsilon = 0.63$, because Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(9) = 17.66$, $p < 0.05$). We further tested the possible visual consequences of the object property change by looking at the absolute change in salience within the boundary of the object across the four change type conditions. Here we also found no main effect of change type on the change in object salience across the cut, $F(1.72, 24.11) < 1$ (Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(5) = 15.70$, $p < 0.01$; therefore the degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity, $\varepsilon = 0.57$). These tests demonstrate not only that the target objects were rarely coincident with perceptually salient locations in the scene, but also that the association between image salience and the object did not vary significantly between the four types of object property change. We can therefore conclude that the perceptual conspicuities of the changes and of the post-change object properties are equivalent.

While the objective salience measure described above suggests that the changes that we made can be considered as visually equivalent between conditions, we can also use subjective measures from the responses of the observers in order to further minimize the possibility of confounds arising from the nature of the changes. One such way of ensuring the subjective perceptual equivalence of the four types of object property change is to measure the 'noticeability' of the change in each video clip. The continuity error question ("Did you notice anything unusual in this video clip?") provides a measure of this noticeability. We considered the measured noticeability of a given change type in two ways. First, we verified that over the course of the experiment no types of (local) change were more noticeable than others. Second, by including the responses to the continuity error question as a covariate in analyses we achieved a high degree of statistical control, factoring out any variance due to any uncontrolled differences in noticeability. We conducted both of these checks in the analyses that follow.

Finally, much prior work on memory for object properties has been subject to a methodological weakness. Effects are typically reported based solely on by-participants analyses while the contribution of particular scenes, images, or films goes unreported. It is important to demonstrate that particular effects are not restricted to a subset of (possibly atypical) experimental items. One solution to this problem, employed in the present paper, is to employ analyses in which items (movies) are entered as a random factor, allowing for examination of the degree to which effects can be generalized across both participants and items (Clark, 1973; see also Baayen, 2008, for a discussion).

Results

Eye movement data

In the present paper analyses are restricted to eye movement measures bearing on object identification. We report two sets of eye-movement data. First, we derived an estimate of the elapsed time following a cut to the first inspection of the target object. These data were analyzed in an "event history" framework. Survival functions were computed for the four types of object property (color, position, type, token). The use of survival (or "hazard") analysis in eye movement research was pioneered by McConkie and colleagues (see Yang & McConkie, 2001, for a discussion). In the present context the technique provides an estimate of the probability for each frame of a given movie that the target will remain un-inspected in the next frame. Second, total inspection time on the target object was treated as the primary dependent variable in linear mixed effects regression (lmer) analyses, using the lme4 package (Bates & Sarkar, 2006) for the R system for statistical computing (R Development Core Team, 2006; see also Baayen, 2008). Participants and scenes were treated as random factors in separate regression analyses for the four groups of participants (divided by four types of local change). The primary reason for adopting this novel analysis strategy, rather than conventional AVOVA, is that it allows for an estimate of the association between success at identifying a change and the pattern of inspection, while controlling for changes in measured noticeability. This was achieved by including performance on the continuity error question in the model. The logic adopted was first to establish a baseline model in which whether or not there was a local change (no change vs. local change) served as a predictor and then to add performance on the continuity error question as an additional predictor, testing the fit of the revised model against the baseline model. Finally, the possibility that whether or not there was a local change and performance on the continuity error question interacted was tested. In all analyses,

	Object Property			
	Colour	Position	Type	Token
Event	118 (78.7%)	114 (76.0%)	120 (80.0%)	126 (84.0%)
Censored	32 (21.3%)	36 (24.0%)	30 (20.0%)	24 (16.0%)

Table 1. Number of event and censored cases for the four types of object property.

the particular set of movies employed was entered as a (dummy) fixed factor. The use of lmer has additional advantages over traditional F1 (by-participants) and F2 (by-items) ANOVAs (Baayen, 2008), in that it allows for a simultaneous evaluation of participant and scene effects and provides a robust statistical test of the independent effects (if any) on the dependent variable of performance on the continuity error question.

Time to first fixation on the target object

The location of each target object in the post-cut static image was defined by a circle fitting the object as closely as possible. A "hit" was counted as the first fixation to fall within the boundary defined by this circle. The time to locate the target was computed in video frames (i.e. units of 40 ms). Time-to-fixate data of this kind demand an "event history" analysis because on a significant proportion of occasions participants will not have fixated the target object when the trial ends (i.e. after 125 frames). Excluding such cases or including them with some arbitrary time to hit value would be inappropriate and may misrepresent the inspection behavior. Survival analysis was developed to deal with data sets taking this form (e.g. estimating survival rates in a drug trial where a proportion of patients will be alive when the study ends). Cases falling into this category in a survival analysis are referred to as "censored". A survival function in the present context provides an estimate of the instantaneous likelihood of fixating the target throughout the whole presentation. Table 1 shows the number of cases where the target was fixated during the trial ("event") and those where the target was not fixated before the trial ended ("censored") for the four types of object property. There was no association between the four types of object property and whether the target object was fixated or censored ($\chi^2(3) = 3.09$, $p > 0.1$). Survival functions for the four types of object property are shown in Figure 3. These plot the probability on each frame of the post-cut scene that at least one fixation will have been on the target object. That is, the function does not simply describe the probability that an object will be receiving fixation during a particular frame, which could include re-fixations of an object, but instead it conveys the probability that the first fixation of an object will occur by this time in the presentation. The data were subjected to a Cox Regression analysis, with whether or not there was a local change (no

change vs. local change), performance on the continuity error question, and scene entered as predictors of time to fixate (see Pannasch, Dornhoefer, Unema, & Velichkovsky, 2001, for a further example of this procedure). Item set was not included since it was a dummy factor (although separate analyses showed that the exclusion did not significantly affect the outcome). The omnibus model significantly predicted time to fixate for each of the four types of object property ($\chi^2(16) = 57.4$; 63.4; 111.3; and 81.4 for color, position, type and token changes respectively, all $p < 0.001$).

When an object changed color during the editorial cut, there was a trend for the object to be looked at sooner than if it did not change and this tendency approached significance ($\chi^2(1) = 3.3$, $p = 0.07$, for the Wald statistic). The difference is apparent in Figure 3. For example, setting the survival probability at 0.3 (i.e. setting the probability that the target will be hit on the next frame at 0.7), this point is not reached until around the 100th frame when the color of the target object remained the same (solid line; Figure 3, top left), whereas it is reached at the 40th frame when the target color changed after the cut (dotted line; Figure 3, top left). Although only marginally significant, it appears that the likelihood of hitting the target in any given frame was slightly higher when the color of the target object changed after the cut. There was a significant tendency for objects to be looked at sooner if they changed to a different type after the cut ($\chi^2(1) = 5.56$, $p = 0.02$), and this was also the case if the object underwent a token change ($\chi^2(1) = 4.7$, $p = 0.03$). As is clear from Figure 3, a position change produced no difference in time to hit the (displaced) target compared to when the object did not change position ($\chi^2(1) < 1$). In marked contrast to the other conditions, for position questions the survival functions for no change and local change conditions are indistinguishable and approach an asymptote well above zero.

In summary, independent of possible changes in the absolute noticeability of a given change, observers were more likely to look earlier at an object if it had undergone a color, type or token change during the cut, than if it had not changed. This result argues in favor of some peripheral detection of the change, resulting in an eye movement towards the object sooner than would otherwise occur in the absence of change. The pattern of results was different for changes in position. With the noticeability of the change controlled, observers were no more
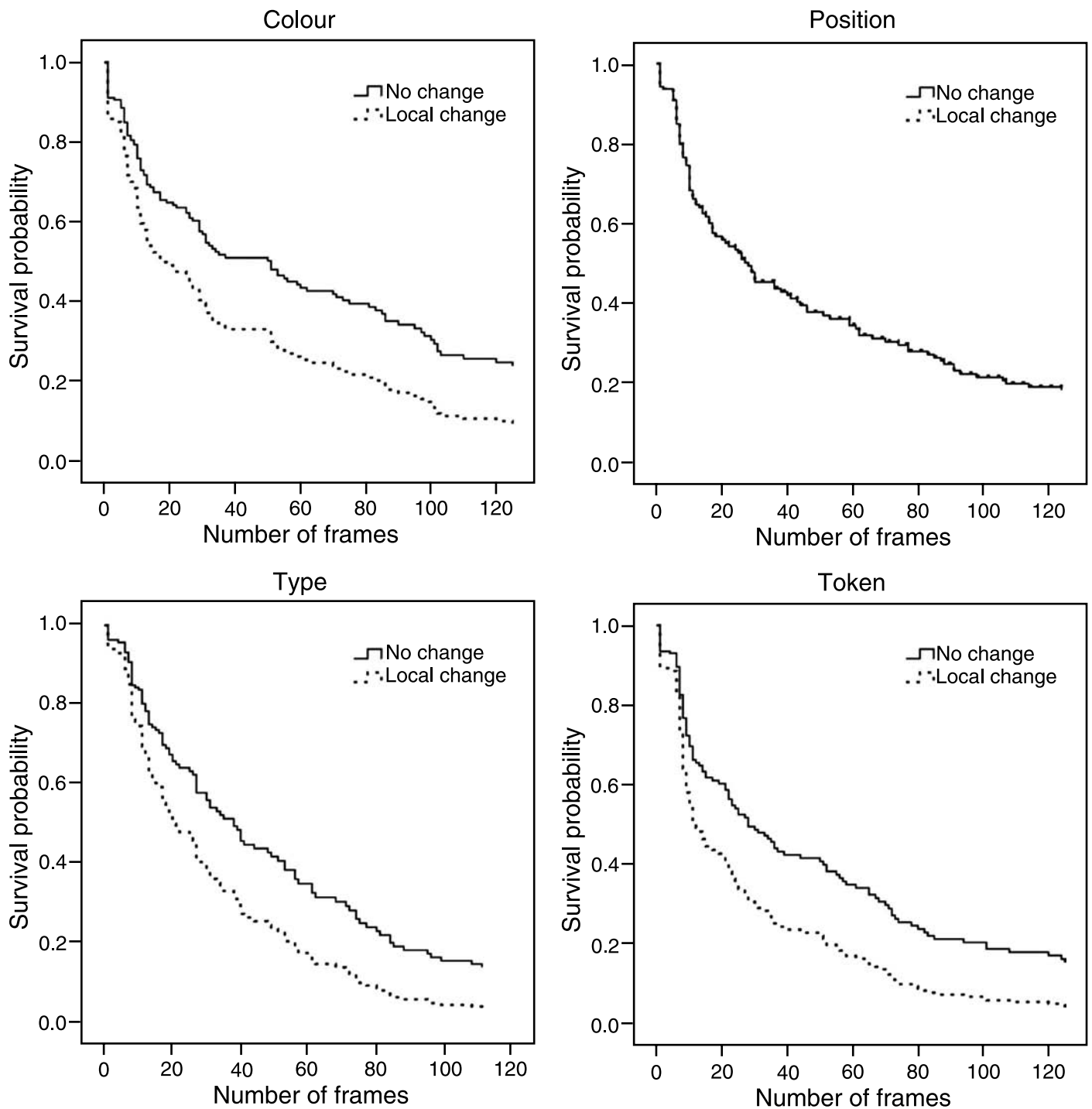
Figure 3. Survival functions for the four types of object property. Solid lines show no change and dotted lines show local change conditions.

likely to fixate the object sooner if all that had changed was its position in the scene after the cut than if it had remained in the same (relative) place. It should be borne in mind, of course, that the post-cut position of the target object in screen coordinates changed quite radically in all conditions. Extending the above logic, we could argue that there was no peripheral detection of a change in position following the post-cut change in represented viewpoint.

Total duration on the target object

Duration in this context was defined as the cumulative number of frames where fixations fell within the boundary of a defined object during a trial. As in the previous section, here we are concerned with comparing trials in which the target object was not changed during the editorial cut, to trial in which it underwent one of the four possible local changes. Separate lmer analyses were

Figure 4. Mean number of frames comprising the total duration on the target object following the cut (with error bars ±1 SE). The four types of object property are shown separately for no change and local change conditions.

conducted for the four types of object property. Figure 4 shows the relevant data, which are contingent on the target actually being fixated. "No hits" were treated as missing data in the analyses, the percentage being roughly the same in each condition (see Table 2).

Table 3 shows the summary results from the four lmer analyses. First we conducted a baseline analysis of the data with whether or not there was a local change (no change vs. local change) and version (three test sets based on a Latin-square design) as the only fixed factors. Performance on the continuity error question (CEQ) was then added to the baseline model and the revised model compared with the baseline. This allowed us to determine whether including the observers' responses to the continuity error question (which is a measure of the noticeability of the local change) improved the fit of the model. This also allows us to effectively account for any influence of the noticeability of a particular local change upon how long objects were fixated. Responses on the continuity error question were marked correct when participants answered "no" to the question "Did you notice anything unusual in this video clip?" in the no change condition and "yes" in the local change condition.

The baseline analysis of the data for the color change condition showed an approximately four frame difference in viewing time following a color change, which was not significant. When performance in the continuity error question was added to the model, the fit of the model improved significantly ($\chi^2(1) = 15.6$, $p < 0.001$). A correct response on the continuity error question (i.e. noticing something "unusual" after a color change) was associated with fixation times about 960 ms longer (i.e. the unstandardized regression coefficient value B multiplied by frame duration of 40 ms, Table 3). Whether or not there was a local color change also had a significant effect on viewing time. Taking continuity error question performance into account in the model, when there was a color change to an object participants looked at it for about 660 ms longer (Table 3). There was also a significant interaction between whether or not there was a local change and how participants performed in the continuity error question. Sub-analyses showed a significant association between longer inspection time and a higher tendency to notice something "unusual" when an object's color changed, but not when it remained the same (see Table 4).

Following a type change there was no significant increase in viewing time for the baseline model. When performance in the continuity error question was added, the model fit improved significantly ($\chi^2(1) = 4.12$, $p < 0.05$). There was again a direct effect of continuity error question performance showing an association between correct response and longer fixation times (in the order of 440 ms, Table 3). Further, when there was a local change to the object type, fixation times differed significantly (by about 320 ms, Table 3). There was no interaction between the occurrence of type changes and performance on the continuity error question.

For token changes, the baseline model showed no significant effect of the occurrence of this type of local change. However, once again, adding continuity error question performance to the model improved its fit significantly ($\chi^2(1) = 9.21$, $p < 0.01$). As in the case of color and type changes described above, noticing something unusual was associated with longer fixation times (about 360 ms longer, Table 3). Furthermore, when continuity error question performance was included in

	Colour		Position	
	No Change	Local Change	No Change	Local Change
Hits	56	62	60	54
No Hits	19	13	15	21
	Type		Token	
	No Change	Local Change	No Change	Local Change
Hits	58	62	60	66
No Hits	17	13	15	9

Table 2. Number of hits and no hits in each of the four types of object property for movies in which the pre- and post-cut scenes matched (no change) and for scenes where the property of object changed after the cut (local change).

| | | Baseline | Baseline + CEQ | | |
		NL	NL	CEQ	NL*CEQ
Colour	B	−4.35	−16.56	23.92	−18.68
	SE	2.81	3.94	5.40	9.13
	t	−1.55	−4.20***	4.43***	−2.05*
Position	B	2.27	9.72	−3.22	−7.64
	SE	2.36	6.81	4.10	8.52
	t	0.96	1.43	−0.78	−0.90
Type	B	−4.53	−8.34	11.02	−11.49
	SE	2.52	3.12	4.20	6.97
	t	−1.80	−2.68**	2.62**	−1.65
Token	B	−2.96	−8.87	9.10	−3.65
	SE	2.07	2.80	3.01	7.49
	t	−1.43	−3.17**	3.02*	−0.49

Table 3. Results from lmer analyses for the four types of object property. *Note*: $*p < 0.05$; $**p < 0.01$; $***p < 0.001$. NL = No change vs Local change; CEQ = Continuity Error Question.

the model, the occurrence of a local change to the token of an object resulted in a change in fixation time in the order of about 320 ms (Table 3). There was no interaction between the occurrence of token changes and performance in the continuity error question.

Finally, for position changes, the data revealed a completely different pattern from the other three object properties. Globally, participants spent about the same amount of time inspecting the target object as in other conditions, but there was no effect of whether or not the object's position changed in the baseline model. Adding continuity error question performance did not improve the model fit ($\chi^2(1) < 1$). That is, noticing something unusual was not associated with any changes in viewing time. Nor did viewing time change depending upon whether the position of an object changed or not, even when the noticeability was accounted for in the model. We also ran a version of these analyses for position changes where the original location of the object (in world-centered frame of reference) was used to code the fixation behavior. In these analyses, we also found no effects of noticeability or whether or not the object position changed.

In summary, observers fixated on the target object longer following changes in color, type and token information and better performance in the continuity error questions was associated with longer fixation time. In contrast, changes in the position of a defined object had no association with the time spent fixating it, again showing a different pattern from other types of object information.

Behavioral data

Performance was measured on each of the four types of object property (color, position, type, token) for three post-pan conditions (no change, global change, local change). The four types of local change made to objects were tested in four separate analyses. Although it is possible to conduct lmer analyses with a dichotomous dependent variable, the technique is not well-suited to compare performance against chance. For the analyses in this section, being able to compare performance to chance is important because we wish to use this approach to consider whether participants select correct answers significantly more frequently than would be expected if they were guessing from the available options. For this reason, the behavioral data were analyzed using separate mixed-design ANOVAs, with participants and movies as random factors. Values of the F statistic derived from by-participants (F1) and by-movies (F2) analyses are reported separately. Since we are a priori interested in whether different memory processes operate for the four types of object property, post-hoc analyses were carried out on these comparisons whether or not significant interactions were found.

Information extraction and integration

Figure 5 shows performance on each of the four types of object property (color, position, type, token) for the two post-pan control conditions (no change vs. global change). The data are collapsed across the four groups of participants who otherwise experienced different types of change. Overall, performance was significantly better when nothing changed in the post-cut scene than when everything changed

| | Colour | |
	No change	Local change
B	5.14	28.02
SE	7.74	5.74
t	0.66	4.88***

Table 4. Comparison between no change and local change conditions for colour. *Note*: $***p < 0.001$.
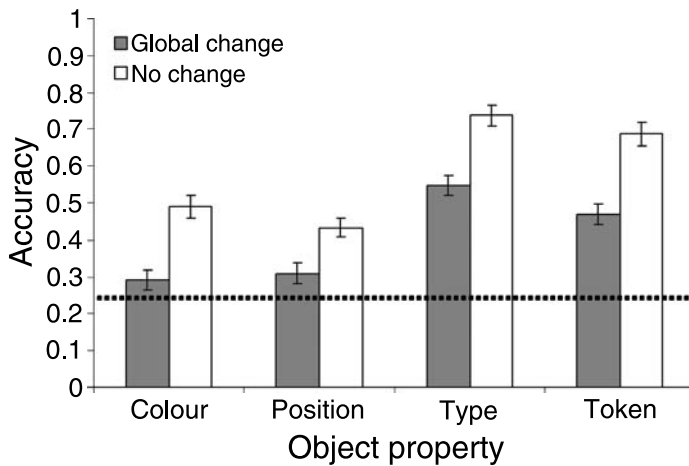
Figure 5. Performance in each of the four types of object property for movies in which the pre- and post-cut sequences matched (no change) and for movies in which the post-cut image was of an entirely different scene from the pre-cut panned sequence (global change). Data are collapsed across the four groups of participants. The dotted line shows the chance level.

($F_1(1, 48) = 65.91, p < 0.001, F_2(1, 12) = 31.48, p < 0.001$), a result that is likely to be due to the additional viewing time in the post-cut scene on the target objects. Performance differed significantly for the four types of object property ($F_1(3, 144) = 65.98, p < 0.001, F_2(3, 36) = 13.06, p < 0.001$), but there was no significant interaction between the type of object property and the two post-pan control conditions ($F_1(3, 144) = 1.24, p > 0.1, F_2(3, 36) < 1$). To examine the memory accuracy for the four object properties in the two post-pan control conditions, performance in each condition was compared to chance. The analyses revealed that performance was not reliably above chance for the color and position questions in the global change condition, whereas performance in all the other conditions was above chance (see Table 5).

In summary, when the target object was visible during the pan but not visible in the post-cut scene, performance was above chance for type and token information but not reliably better than chance for position and color information. This outcome emphasizes the importance of by-items analyses as it is not the conclusion that would have been reached on the basis of by-participants analyses alone. The by-participants analyses for color and position information showed performance to be above chance. However, the by-items analyses were non-significant for color and position, showing that the apparent effects in the by-participants analyses were only for a sub-set (possibly as small as one) of movies. It should be noted that the global change condition provides a conservative estimate of the information encoded from the pan, as the novel post-cut scene effectively masks any memory encoded from the pan. The fact that two types of identity information were successfully retained from the pan but position and color were not, suggests that there are

qualitative differences in the manner in which these object properties are encoded and retained. This finding suggests that both type and token object identity information was retained across the post-cut scene, regardless of whether or not it contained the target object, and was still available at the time of the memory test. The outcome extends the findings of Wallis and Bülthoff (2000), and indicates that the coding of object identity information appears to be more stable across changes in viewpoint.

For color and position information the situation is less clear-cut. The fact that performance was only reliably above chance when the object was visible before and after the cut, and was not reliably above chance when the object was only present in the panned sequence, means that definitive statements about whether these sources of information are integrated across changes in viewpoint cannot be made. It may be that position and color information were not encoded during the pan or were forgotten by the time the test was administered, or all information for these two object properties could have been acquired from the post-cut static scene alone. Alternatively, pre- and post-cut information may have been integrated, but the pre-cut information simply not retained for later recognition when the post-cut scene did not contain the target object. We will return to the question of whether color and position information from the panned sequence appears to be encoded in the Discussion.

It should also be noted that when viewing the movie sequences with "no change" at the cut, recognition performance varied between the four types of object property. Participants' memory performance of which particular object had been present (i.e. type and token) was much better than that of properties such as color or position. This differential performance (whether or not it is the result of a failure to encode position and color during the pan) provides further evidence that different object properties behave differently in memory and is consistent with previous suggestions of differential encoding and retention of each object characteristic (Tatler et al., 2005).

	No Change	Global Change
Colour	$F_1(1, 48) = 59.25***$ $F_2(1, 12) = 38.80***$	$F_1(1, 48) = 3.60^+$ $F_2(1, 12) = 1.47$
Position	$F_1(1, 48) = 54.50***$ $F_2(1, 12) = 10.10**$	$F_1(1, 48) = 4.29*$ $F_2(1, 12) = 1.88$
Type	$F_1(1, 48) = 394.74***$ $F_2(1, 12) = 129.44***$	$F_1(1, 48) = 120.93***$ $F_2(1, 12) = 30.88***$
Token	$F_1(1, 48) = 231.91***$ $F_2(1, 12) = 45.64***$	$F_1(1, 48) = 68.06***$ $F_2(1, 12) = 10.27**$

Table 5. F_1 and F_2 values when performance in each of the four types of object property was compared to chance for no change and global change conditions. Note: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; $^+$marginally significant at $p = 0.06$.
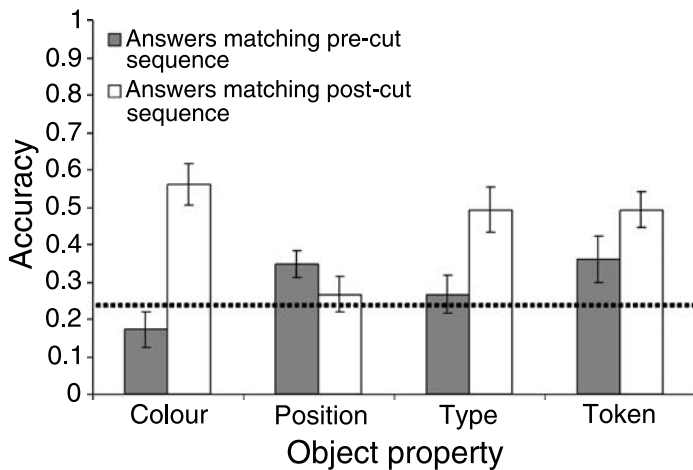
Figure 6. Performance on each of the four types of object property. The figure shows the proportion of responses matching the pre-cut or post-cut version of the object when a single object property was changed. The dotted line shows the chance level.

Memory consequences of changes to objects

In this section we examine performance on the four types of object property as a function of the particular local change made. In this case, there was not only a change in represented viewpoint, but also a change in one particular property of a target object in the pre- and post-cut scenes (e.g. a bag visible in the pan was red prior to the cut but gray in the post-cut static scene). The relevant data are shown in Figure 6. Overall, observers appeared to base their responses on the properties of the object visible after the cut (F1(1, 48) = 17.42, p < 0.001, F2(1, 12) = 5.50, p = 0.03). The interaction between the particular local change made and whether the response was based on pre- or post-cut scenes was not reliable, but was significant for a subset of scenes (F1(3, 48) = 5.96, p = 0.002, F2(3, 36) = 2.05, p > 0.1). However, since we have a priori grounds for carrying out the analyses, performance based on pre- and post-cut scenes were compared separately for the four types of local change.

We return at this point to the possible role played by uncontrolled differences in "noticeability" between the four types of local change. It could be argued, for example, that performance in the position task was simply a consequence of the way the films were constructed. One possibility is that position changes were in some way less noticeable compared with other types of change. However, this appears not to be the case (Table 6). There was, in fact, only a hint of a difference in continuity error performance between the four conditions, restricted to the by-scenes analysis (F1(3, 56) = 1.48, p > 0.1, F2(3, 36) = 2.92, p = 0.05). It is also possible, but unlikely, that differences in target-foil difficulty between conditions were not random, but acted somehow as a source of systematic bias. We have no data to address this particular concern directly, apart from pointing to the care taken in constructing the materials.

Following the logic adopted in the eye movement analyses it is possible to assess the contribution of noticeability in each experiment by entering the average value on the continuity error question for each of the 15 movies as a covariate in the by-items analyses. As reported above, the interaction between the particular local change made and whether the response was based on pre- or post-cut scenes was not significant by scenes, albeit not zero. If it is the case that this interaction effect derived primarily from uncontrolled differences in how noticeable a given change was, then it should disappear, or radically reduce, when the noticeability index is added as a covariate in the by-items analyses. This was not the case: treating the four types of local change as a between-items factor, the (non-significant) value of F2, after controlling for noticeability, was virtually identical to the F2 value without a covariate (without noticeability as a covariate, F2(3, 48) = 2.01; with noticeability as a covariate, F2(3, 47) = 1.91). We conclude that it is not the case that the obtained modest interaction arises as an artifact of differential noticeability. Even treating the four types of change as a within-items factor and using an overall average noticeability, rather than by-condition averages, gave a similar result (with noticeability as a covariate, F2(3, 33) = 1.18).

Performance based on pre- and post-cut scenes in the four local change conditions is reported below. As an additional measure of any response biases, in each condition we considered whether the participants were more likely to select as their response either the pre- or post-cut version of the object property than would be expected by chance. In each case, we report results from by-items analyses with and without the noticeability index included (see Table 7).

When the color of the target object changed during a cut, observers were more likely to base their answers on the color of the object visible after the cut (F1(1, 12) = 27.13, p < 0.001, F2(1, 12) = 23.36, p < 0.001). That is, if a handbag changed color from red to gray, observers were likely to answer that the bag had been gray. Moreover, participants selected the post-cut color of the object significantly more often than would be expected by chance. In contrast, they were no more likely to select the pre-cut color than would be expected by chance. With variance due to noticeability removed the pattern remained the same (F2(1, 11) = 6.77, p = 0.03), suggesting that the results were not influenced by how frequently the change was noticed by the observers.

	Local Change
Colour	0.17 (0.06)
Position	0.15 (0.06)
Type	0.33 (0.08)
Token	0.24 (0.07)

Table 6. Overall noticeability rates in the local change condition (standard errors in parentheses).

	Pre-cut	Post-cut
Colour	$F_1 (1, 14) = 2.63$	$F_1 (1, 14) = 30.76^{***}$
	$F_2 (1, 12) = 4.72$	$F_2 (1, 12) = 23.50^{***}$
	VR $F_2 (1, 11) = 2.76$	VR $F_2 (1, 11) = 6.02^{*}$
Position	$F_1 (1, 14) = 7.08$	$F_1 (1, 14) = 0.13$
	$F_2 (1, 12) = 1.05$	$F_2 (1, 12) = 0.04$
	VR $F_2 (1, 11) = 1.07$	VR $F_2 (1, 11) = 0.08$
Type	$F_1 (1, 14) = 0.11$	$F_1 (1, 14) = 15.75^{***}$
	$F_2 (1, 12) = 0.06$	$F_2 (1, 12) = 8.03^{*}$
	VR $F_2 (1, 11) = 0.14$	VR $F_2 (1, 11) = 0.13$
Token	$F_1 (1, 14) = 3.11$	$F_1 (1, 14) = 26.49^{***}$
	$F_2 (1, 12) = 1.94$	$F_2 (1, 12) = 7.40^{*}$
	VR $F_2 (1, 11) = 2.05$	VR $F_2 (1, 11) = 0.78$

Table 7. F1, F2 and Variance Removed F2 values when performance in each of the four local change conditions was compared to chance. *Note*: $^{*}p < 0.05$; $^{***}p < 0.001$.

In type and token change conditions observers were equally likely to base their answers on pre- or post-cut scenes (type change, $F_1(1, 12) = 6.15$, $p = 0.03$, $F_2(1, 12) = 2.37$, $p > 0.1$; token change, $F_1(1, 12) = 2.02$, $p > 0.1$, $F_2(1, 12) < 1$). Removal of variance due to noticeability did not change this pattern (type change, $F_2(1, 11) < 1$; token change, $F_2(1, 11) < 1$). The results also showed that participants selected the post-cut identity of the object more frequently than would be expected by chance. In contrast, participants were no more likely to select the pre-cut object identity than would be expected by chance. However, with variance due to noticeability taken out, neither the pre- nor the post-cut object identity was selected more frequently than expected by chance, confirming that variations in noticeability almost certainly explain differences in performance in these cases.

When the object's position was changed during the cut, responses were no more likely to be based on the position of the object in the post-cut static scene than on its position in the pre-cut (panned) sequence ($F_1(1, 12) = 2.25$, $p > 0.1$, $F_2(1, 12) < 1$). Removal of variance due to noticeability did not affect the outcome ($F_2 < 1$). Furthermore, participants were no more likely to select either pre- or post-cut positions than would be expected by chance, and removing variance due to noticeability did not alter this pattern. This outcome is consistent with earlier studies (e.g. Levin & Simons, 1997; Wallis & Bülthoff, 2000) indicating difficulty coding position information in dynamic scenes.

Discussion

The central aim of the present experiment was to examine inspection behavior and memory around the time of a cut in moving images. Specifically, we measured eye movements and object memory performance to infer the types of information that are encoded across cuts between a panned sequence and the same scene from a new viewpoint. By introducing changes to particular object properties across a cut, we could consider the differential sensitivity to these changes in eye movement and memory performance. Sensitivity was assessed by comparing inspection behavior and memory when a single object property was changed to when either nothing changed across the cut, or the cut was to an entirely different scene. In this dynamic setting we see evidence for information extraction and retention from panned movie sequences. We find oculomotor behavioral consequences of changing the properties of an object during a cut. In general, changed objects are looked at sooner than they would be if the relevant property had not changed and, once fixated, they are looked at for longer than if they had not changed.

However, consistent with our general hypothesis regarding the sub-structure of visual representations, the above general pattern was only present for color and two types of identity information and not for position information. This suggests that stored information about the color and identity of objects is sufficient to result in sensitivity to changes in these object properties. Similar evidence for stored object information across a cut can be found in the object memory performance data. Here performance is better when the panned and post-cut scenes are the same compared to when the pre- and post-cut scenes differ. These findings are in line with previous research from static scene viewing using change detection paradigms (for a comprehensive summary of early work, see Henderson & Hollingworth, 1999) or recognition performance (e.g. Tatler et al., 2005). In contrast, consistent with our second specific hypothesis, across all measures the representation of position information is quantitatively distinct, an outcome which does not arise from artifactual changes in the inherent noticeability of particular local changes. Of course, position may be considered to be a qualitatively different type of object information than color or identity. Position describes the relationship of the object to some external frame of reference, whereas the other types of information tested are to some degree independent of external factors and as such are internal to the object.

The present data can be used to comment on the sub-structure of visual representations, a topic on which the existing literature is less clear-cut (Angelone et al., 2003; Levin & Simons, 1997; Wallis & Bülthoff, 2000). For static scenes, particular object properties are encoded and retained to differing degrees and over differing timescales (e.g. Tatler et al., 2003, 2005). However, there is some disagreement about how particular types of information are encoded. For example, it is rather unclear whether identity and color information accumulate with viewing time or not (contrast Tatler et al., 2005, with Hollingworth & Henderson, 2002 and Melcher, 2006). We will here for convenience distinguish position and the other tested sources of information as 'external' and 'internal' object

properties respectively. What we refer to as 'internal' features of objects should not be confused with intrinsic object properties. The former refers to features that are not reliant on relationships with external factors such as other objects or frames of reference. The latter refers to features of objects that are fundamental to its identity. Our discussion does not consider intrinsic object properties. We will use our data to argue for sub-structure of visual representations in the context of our experiment, where internal and external object properties are represented in distinct ways. More subtle differences in the representation of different types of internal object property are also evident.

Noticeability vs. behavioral consequences of change

While it should be remembered that our experiment was not a change detection study per se, we did ask observers to indicate whether they saw anything unusual (in the form of a continuity error) after each movie sequence. This measure was used to estimate how noticeable our four types of local change were to observers. There were no statistically significant differences in the noticeability of changes to color, position, type or token information (see Table 6). In contrast, we found behavioral differences in the consequences of these four types of local change both in terms of oculomotor inspection and recognition memory performance. This result hints at the possibility of some degree of dissociation between explicit awareness of change (as indexed by noticeability) and (implicit) behavioral consequences of change. The results for noticeability show very low rates of reporting something unusual in the movies when local changes were made. In general, observers neither noticed nor showed behavioral responses to changes to the position of objects in movies. In contrast, despite rarely noticing changes to the other object properties, we did find differences to the inspection behavior and to recognition memory performance when changes were made to the color or identity (either type or token) of objects. Because the present study was not an explicit change detection study we can only speculate, but the results could be interpreted as offering support for the notion that changes can have measurable behavioral consequences even in the absence of explicit detection (e.g. Fernandez-Duque & Thornton, 2000; Silverman & Mack, 2001).

Representing internal object properties

Consider first oculomotor measures of the observers' sensitivity to change. Broadly speaking, the same patterns of inspection are evident for color, type and token changes made to objects in the scenes. In all cases, if the property is changed during the cut to a new camera angle, observers look at the changed object sooner and for longer than if it had not changed. The fact that the change influences subsequent viewing behavior suggests that the representation of the pre-cut information is sufficient for some form of detection (be it implicit or explicit) of an inconsistency between the object before and after the cut. Since there are no quantifiable differences in the influence of changes for these three 'internal' properties, the eye movement record does not allow us to distinguish between the representations underlying these changes in behavior. But while oculomotor measures cannot distinguish between color, type and token information, recognition performance in the memory questions can.

Compare first memory for an object seen only in the panned sequence (where the post-cut scene was globally different from the panned sequence) and contrast that with memory for objects present (without any changes) during both the panned sequence and the post-cut static scene (Figure 5). For both type and token information there is evidence of encoding and retention from the panned sequence as well as post-cut scene. We also found no response selection bias between pre- and post-cut scenes for these two types of identity information when they were changed during the editorial cut (Figure 6). This may suggest that identity information from both before and after the cut is encoded and retained, with no bias to select either of these stored states of the object. However, it should be noted that only the post-cut object properties were actually reported at an above-chance level and even this difference disappeared once the noticeability of the change was included in the model.

With regard to memory for object color, recognition performance showed a different pattern. Color information was not remembered when only seen in the panned sequence, but if seen both in the pan and the post-cut scene memory for color reached levels above chance (Figure 5). There are also differences between color and the other two types of internal object property when local object change was made. Participants based their answers to color questions very strongly on the color of the object after the cut, and only the post-cut object color was selected more frequently than would be expected by chance (Figure 6). While this trend was evident for type and token questions, the bias was far less pronounced and was largely explained by variations in noticeability. It is possible that color information may have been encoded during the pan but subsequently overwritten by that encoded during the post-cut scene. Overwriting of represented information is consistent with suggestions from some change detection studies (e.g. Simons, 2000) and with the suggested retention of visually rich information under more natural conditions (Tatler, 2001). It is also possible that information from before and after the cut was retained and represented (e.g. Mitroff, Simons, & Levin, 2004; Simons, 2000), but that participants chose to base their answers on the most recently viewed version of the object. On the other hand, it will be recalled that memory performance for object color was actually at chance when the object was only visible during the pan and was not present in the post-cut scene. From this it

might be argued that color information was simply not encoded during the pan. However, the analyses of oculomtor behavior provide a strong indication that color information was, in fact, extracted from the pre-cut panned sequence (Figures 3 and 4).

These differences in recognition performance between color and both type and token information may arise for a number of reasons. The combination of not being able to remember color information for objects only visible in the panned sequence and the finding that when the color changed across the cut, responses were based heavily on the post-cut color, suggests that we are only able to access information about color for very recently viewed objects. In contrast, for type and token information we appear to be able to access information about objects only visible in the pan, and responses in post-trial questions were less heavily biased to the post-cut object properties. These differences may suggest differential biases in retrieval, differential accessibility or even differential encoding and representation of color compared to type and token information. However, our data cannot distinguish between these possible explanations of the results. It will be for future experiments to tease apart these possible interpretations. It should be noted, however, that the pattern of differences can only satisfactorily be explored using dynamic, rather than static, images.

Representing external object properties

In all measures, object position information showed very different patterns from other types of information tested. Position information was not retained from the panned sequence if the post-cut scene did not contain the object (Figure 5), and when the position of an object in a scene changed after a cut, participants were not biased towards the post-change location of the object. Indeed, there was a non-significant trend in the opposite direction (Figure 6). The oculomotor consequences of changing the location of an object in a scene were also qualitatively distinct from those resulting from changes to any other object property tested. When the position of an object changed, observers were no more likely to look at it sooner than if it had not changed (Figure 3), and it was not looked at for any longer than if its position had not changed when the object in its changed location was examined (Figure 4). This lack of oculomotor consequences of change is in marked contrast to the results found for changes to internal object properties. However, this does not indicate that position information is never encoded. When an object did not change in any way across the cut, performance in the position questions was significantly above chance (Figure 5). Hence what we see in the trials in which object position was changed during the cut is a specific failure to deal with a change in the position information. The outcome goes some way towards discriminating between whether position information was encoded in the pan but not remembered, or was never

encoded in the first place. If it were the case that position information was only encoded during the post-cut static scene, then changing the position of the target object during the cut should not be disruptive: observers would simply encode the post-cut information from the static scene and provide an answer based on the position of the target object after the cut. The results do not show this pattern. Performance drops to chance when object position is changed, suggesting that information from both before and after the cut is encoded but the conflict between these sources of information cannot easily be resolved. Certainly, this result suggests that position information is represented in a rather different manner to the other types of information tested. This is consistent with previous research using static scenes which have suggested that position information may be represented in a rather different manner from other object information (e.g. Aginsky & Tarr, 2000; Rensink, 2000; Tatler et al., 2003, 2005). As such, we can use the present data to extend the results from static scene viewing to the more dynamic setting of watching movie sequences.

The recognition memory performance together with the lack of oculomotor consequences of position change suggests that our observers had great difficulty representing the position of the objects accurately. Why might this be the case? Paucity of position representation in dynamic scenes is in some ways at odds with a range of previous studies regarding vision and spatial representation. In experiments using static scenes, memory performance on questions relating to position (e.g. Melcher, 2006; Tatler et al., 2003, 2005) and detection of object position changes (e.g. Simons, 1996) are typically good. Similarly, there is ample evidence from studies of real world behavior to suggest that in dynamic environments position information is represented well enough to target eye movements to remembered object positions (Ballard et al., 1992, 1995; Land, Mennie, & Rusted, 1999). If based solely on results from static scenes and natural behavior, we should also expect good memory coding in moving images. However, our current finding of poor position coding in dynamic scenes suggests that there are problems specific to movies. Moving images call on the ability to integrate position information across a large and abrupt change in viewpoint, something that is very unlikely to occur in a natural setting. Indexing the position of an object within the scene requires knowledge of how the object relates to the environment it is in and possibly to the local context of other landmarks present around it (e.g. in the "pattern-centric" co-ordinates suggested by Wade & Swanston, 1996). That is, when viewing static pictures the co-ordinate system in which memory-guided saccades are encoded is likely to be the picture frame itself or 'landmark' objects (Kennedy, Brooks, Flynn, & Prophet, 2002). In the case of static images, disruption to the relative spatial relationships of objects in context reduces the advantage in recognition performance for objects presented in the same scene location (Hollingworth, 2007), suggest-

ing that visual representations of objects are bound to scene locations and object position must be defined in relation to a larger (static) scene representation. Coding of this form is very likely to fail in the case of dynamic images, but cannot readily be replaced by the kind of egocentric coding employed in normal interaction with the environment. In a movie, the physical location (in screen co-ordinates) of an object may be unchanged even if it has, nonetheless, radically changed its position relative to other objects. Equally, the reverse may be true. It follows that the coding of position in movies may be independent of, and secondary to, the coding of internal object properties. Further, our data suggest that it is relatively difficult in absolute terms. We may be more sensitive to internal object changes in dynamic scenes because these are properties unlikely to change in the real world. Conversely, the position of an object in a dynamic environment can change either as a result of ego-motion of the observer or manipulation by the observer. As such, we may be more tolerant of shifts in object position (Tatler et al., 2005). In a movie, when an abrupt change in viewpoint occurs, this disrupts the whole scene. The viewer must deal with the sudden movement of all of the elements present and form a coherent representation of the scene, integrating if possible what came before with what came after the cut. Given the need to resolve a large number of changes resulting from the new viewpoint, it is less surprising that observers might fail to notice that a single object has not shifted to an equal extent to the other elements in the scene.

Conclusion

The current study examined how different types of object property are represented in moving images con-taining viewpoint changes. We introduced changes to different object properties across a cut and examined eye movement behavior and recognition memory. Our data demonstrate that internal object properties such as color or identity are represented across viewpoint changes in moving images, although memory is generally biased towards information that is most recent. The representation of position information, in contrast, showed a different pattern of results in both eye movement behavior and recognition memory performance. When viewing moving images position information is not represented in the same way as other object properties, and there is little evidence of the effect of position change on memory and eye movement. Thus the current study demonstrates the importance of considering sub-structure of visual representations, and the results have implications for the ways in which observers construct and maintain a coherent representation of the complex visual environment from viewing dynamic scenes containing viewpoint changes.

Appendix A

Descriptions of questions used

A question regarding a continuity error:

1. Did you notice anything unusual in this video clip?

Four-alternative forced choice questions:

1. Which of the following best matches an object in the scene?

Video Descriptions

Scene	Object Colour	Original Target	Colour Change	Position Change	Type Change	Token Change
Bookshop	Green	Ring binder	Blue	Right	Notebook	Folder
Café A	Blue	Teapot	Pink	Left	Biscuit jar	Different teapot
Café B	Blue	Astray	green	Right	Sugar bowl	Different ashtray
Café C	Brown	Jug	Cream	Right	Coffee jar	Different jug
Common Room A	Silver	Electric kettle	White	Left	Toaster	Hob kettle
Common Room B	Black	Mug	White	Left	Teapot	Cup
Foyer	Pink	Stapler	Blue	Left	Scissors	Different stapler
Gym	Brown	Towel	Blue	Right	Gym shoes	Tea towel
IT Suite	Black	Penholder	Blue	Right	Telephone	Different penholder
Kitchen	Green	Casserole dish	Grey	Right	Colander	Sauce pan
Lecture Room	Red	Handbag	Grey	Right	Hat	Gift bag
Library	Red	Screwdriver	Blue	Left	Umbrella	Pliers
Office	Black	Desk lamp	White	Left	Thermos	Anglepoise lamp
Union A	Pink	Purse	White	Left	Gloves	Different purse
Union B	White	Pool ball	Red	Right	Cigarette packet	Golf ball

Table B1. Short descriptions of movie scenarios used. *Note*: The second column lists the object colour common to all change type (except for the colour change condition) and the original target.

2. Which of the following best matches the correct position of "the object name" in the scene?
3. Which of the following best matches the correct color of "the object name" in the scene?
4. Which of the following best matches "the object name" in the scene?

Appendix B

Table B1.

Acknowledgments

Commercial relationships: none.
Corresponding author: Yoriko Hirose.
Email: y.hirose@dundee.ac.uk.
Address: School of Psychology, University of Dundee, Dundee, DD1 4HN, Scotland, UK.

References

Aginsky, V., & Tarr, M. J. (2000). How are different properties of a scene encoded in visual memory? *Visual Cognition, 7,* 147–162.

Angelone, B. L., Levin, D. T., & Simons, D. J. (2003). The relationship between change detection and recognition of centrally attended objects in motion pictures. *Perception, 32,* 947–962. [PubMed]

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics.* Cambridge: Cambridge University Press.

Ballard, D. H., Hayhoe, M. M., Li, F., Whitehead, S. D., Frisby, J. P., Taylor, J. G., et al. (1992). Hand eye coordination during sequential tasks. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 337,* 331–339. [PubMed]

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7,* 66–80.

Bates, D., & Sarkar, D. (2006). *lme4: Linear mixed-effect models using S4 classes.* R package version 0.995-2, http://lme4.r-forge.r-project.org.

Bordwell, D., & Thompson, K. (2001). *Film art: An introduction.* New York, USA: McGraw-Hill.

Buratto, L. G., Matthews, W. J., & Lamberts, K. (2009). When are moving images remembered better? Study-test congruence and the dynamic superiority effect. *Quarterly Journal of Experimental Psychology, 62,* 1896–1903. [PubMed]

Carmi, R., & Itti, L. (2006). The role of memory in guiding attention during natural vision. *Journal of Vision, 6*(9):4, 898–914, http://journalofvision.org/6/9/4/, doi:10.1167/6.9.4. [PubMed] [Article]

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12,* 335–359.

Droll, J. A., Hayhoe, M. M., Triesch, J., & Sullivan, B. T. (2005). Task demands control acquisition and storage of visual information. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1416–1438. [PubMed]

Fernandez-Duque, D., & Thornton, I. M. (2000). Change detection without awareness: Do explicit reports underestimate the representation of change in the visual system? *Visual Cognition, 7,* 323–344.

Garsoffky, B., Huff, M., & Schwan, S. (2007). Changing viewpoints during dynamic events. *Perception, 36,* 366–374. [PubMed]

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

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Atkins (Ed.), *Perception: Vancouver studies in cognitive science* (vol. 2, pp. 89–110). New York: Oxford University Press.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50,* 243–271. [PubMed]

Henderson, J. M., & Hollingworth, A. (2003). Eye movements and visual memory: Detecting changes to saccade targets in scenes. *Perception & Psychophysics, 65,* 58–71. [PubMed]

Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 31–47. [PubMed]

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

Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene perception: Memory for things observed. *Perception & Psychophysics, 64,* 882–895. [PubMed]

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40,* 1489–1506. [PubMed]

Kennedy, A., Brooks, R., Flynn, L.-A., & Prophet, C. (2002). The reader's spatial code. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 193–212). Oxford: Elsevier.

Land, M. F., Mennie, N., & Rusted, J. (1999). The role of vision and eye movements in the control of activities of daily living. *Perception, 28,* 1311–1328.

Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin and Review, 4,* 501–506.

Matthews, W. J., Benjamin, C., & Osborne, C. (2007). Memory for moving and static images. *Psychonomic Bulletin and Review, 14,* 989–993. [PubMed]

Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision, 6*(1):2, 8–17, http://journalofvision.org/6/1/2/, doi:10.1167/6.1.2. [PubMed] [Article]

Melcher, D., & Kowler, E. (2001). Visual scene memory and guidance of saccadic eye movements. *Vision Research, 41,* 3597–3611. [PubMed]

Mitroff, S. R., Simons, D. J., & Levin, D. T. (2004). Nothing compares 2 views: Change blindness can occur despite preserved access to the changed information. *Perception & Psychophysics, 66,* 1268–1281. [PubMed]

O'Regan, J. K. (1992). Solving the "real" mysteries of visual-perception: The world as an outside memory. *Canadian Journal of Psychology-Revue Canadienne De Psychologie, 46,* 461–488. [PubMed]

O'Regan, J. K., & Noë (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences, 24,* 939–973. [PubMed]

Pannasch, S., Dornhoefer, S. M., Unema, P. J. A., & Velichkovsky, B. M. (2001). The omnipresent prolongation of visual fixations: Saccades are inhibited by changes in situation and in subject's activity. *Vision Research, 41,* 3345–3351. [PubMed]

R Development Core Team (2006). *R: A language and environment for statistical computing* (Version 2.3.1). R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, http://www.R-project.org.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition, 7,* 17–42.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive change in scenes. *Psychological Science, 8,* 368–373.

Silverman, M., & Mack, A. (2001). Priming from change blindness [Abstract]. *Journal of Vision, 1*(3):13, 13a, http://journalofvision.org/1/3/13/, doi:10.1167/1.3.13.

Simons, D. J. (1996). In sight, out of mind: When object representations fail. *Psychological Science, 7,* 301–305.

Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition, 7,* 1–15.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1,* 261–267.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences, 9,* 16–20. [PubMed]

Smith, T. J., & Henderson, J. M. (2008). Edit blindness: The relationship between attention and global change blindness in dynamic scenes. *Journal of Eye Movement Research, 2,* 1–17. [Article]

Tatler, B. W. (2001). Characterising the visual buffer: Real-world evidence for overwriting early in each fixation. *Perception, 30,* 993–1006. [PubMed]

Tatler, B. W., Gilchrist, I. D., & Land, M. F. (2005). Visual memory for objects in natural scenes: From fixation to object files. *Quarterly Journal of Experimental Psychology, 58A,* 931–960. [PubMed]

Tatler, B. W., Gilchrist, I. D., & Rusted, J. (2003). The time course of abstract visual representation. *Perception, 32,* 579–592. [PubMed]

Tatler, B. W., & Melcher, D. (2007). Pictures in mind: Initial encoding of object properties varies with the realism of the scene stimulus. *Perception, 36,* 1715–1729. [PubMed]

Triesch, J., Ballard, D. H., Hayhoe, M. M., & Sullivan, B. T. (2003). What you see is what you need. *Journal of Vision, 3*(1):9, 86–94, http://journalofvision.org/3/1/9/, doi:10.1167/3.1.9. [PubMed] [Article]

Wade, N. J., & Swanston, M. T. (1996). A general model for the perception of space and motion. *Perception, 25,* 187–194. [PubMed]

Wallis, G., & Bülthoff, H. (2000). What's scene and not seen: Influences of movement and task upon what we see. *Visual Cognition, 7,* 175–190.

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks, 19,* 1395–1407. [PubMed]

Yang, S. N., & McConkie, G. W. (2001). Eye movements during reading: A theory of saccade initiation times. *Vision Research, 41,* 3567–3585. [PubMed]