Psychonomic Bulletin & Review
1997, 4 (4), 501-506

BRIEF REPORTS

Failure to detect changes to attended objects in motion pictures

DANIEL T. LEVIN and DANIEL J. SIMONS
Cornell University, Ithaca, New York

Our intuition that we richly represent the visual details of our environment is illusory. When viewing a scene, we seem to use detailed representations of object properties and interobject relations to achieve a sense of continuity across views. Yet, several recent studies show that human observers fail to detect changes to objects and object properties when localized retinal information signaling a change is masked or eliminated (e.g., by eye movements). However, these studies changed arbitrarily chosen objects which may have been outside the focus of attention. We draw on previous research showing the importance of spatiotemporal information for tracking objects by creating short motion pictures in which objects in both arbitrary locations and the very center of attention were changed. Adult observers failed to notice changes in both cases, even when the sole actor in a scene transformed into another person across an instantaneous change in camera angle (or "cut").

Our immediate experiences of the world seem to include rich and detailed visual information about the locations and properties of objects. If this is true, then we should readily detect changes to objects in our environment. Yet, a number of recent findings show that observers are surprisingly slow and often fail to detect changes to successive views of both natural and artificial scenes (Blackmore, Brelstaff, Nelson, & Troscianko, 1995; Grimes, 1996; McConkie & Currie, 1996; McConkie & Zola, 1979; O'Regan, Rensink, & Clark, 1996; Pashler, 1988; Rensink, O'Regan, & Clark, 1996; Simons, 1996; Tarr & Aginsky, 1996). Recent experiments have used a paradigm in which observers view two rapidly alternating versions of a photographed natural scene that differ by a change to an object or an object part (e.g., Blackmore et al., 1995; Rensink et al., 1996). Observers simply try to detect the changing element. Assuming apparent motion is eliminated or masked by a blank interval (e.g., Rensink et al., 1996) or by the simultaneous appearance of additional objects (O'Regan et al., 1996), observers require a large number of alternations before finally identifying the change. Here we extend this finding by showing that change-detection failures occur not only for objects at some arbitrarily chosen location in a scene, but also for the very object that is the center of attention.

One way of understanding previous change-detection failures is to assume that long detection latencies are generally consistent with a serial search among the objects in the scene which terminates when the search process reaches the changing object. Noticing a change requires successively attending to and encoding each object in the scene. Accordingly, changes to unattended objects will go unnoticed, but property changes will be detected immediately if the changed object becomes the center of attention. This hypothesis is supported by findings that changes to objects rated as more important or central in a scene are detected more quickly (Rensink et al., 1996); they are given higher priority in a limited-capacity visual search. Together with evidence that people cannot visually integrate information across eye movements (Bridgeman & Mayer, 1983; Irwin, Brown, & Sun, 1988; Irwin, Yantis, & Jonides, 1983; Jonides, Irwin, & Yantis, 1983) and that briefly presented pictures are quickly forgotten (Intraub, 1981; Potter, 1976), these findings suggest that only the objects that receive focused attention are retained across views.

If the properties of centrally attended objects are represented more fully and are maintained across fixations, these might be the basis of a perceptually rich betweenview representation. Serial-search accounts often assume that attention is not only necessary, but also sufficient for change detection; changes to an object that is the center of attention will be detected. Although the notion that at-

The authors contributed equally to this report, and authorship order was determined arbitrarily. The contents of this paper were included as part of D.J.S.'s doctoral dissertation. Thanks are extended to Ulric Neisser, Romi Nijhawan, Julie Noland, Kathy Richards, Carter Smith, Michael SpiveyKnowlton, and Arthur Woll, for reading earlier drafts of this report, and to Mark Andrews, Justin Barrett, James Beale, Laura Free, Grant Gutheil, Sabina Larnsfuss, Carole Lunney, Julie Noland, Kathy Richards, and Andrea Rosati, for appearing as actors in the films. D.J.S. was supported by NSF and Jacob K. Javits fellowships. Correspondence may be directed to either author at the following addresses: D. T. Levin, Department of Psychology, Kent State University, Kent, OH 44242 (e-mail: dlevin @kent.edu) and D. J. Simons, Department of Psychology, Harvard University, Cambridge, MA 02138 (email: dsimons@wjh.harvard.edu).

tended objects will be more fully represented has some empirical and intuitive appeal, even objects that are the center of interest may require extensive, effortful processing to be represented precisely. Here we examine the role of attention in change detection and consider the possibility that even if attending to an object is necessary for change detection, it may not be sufficient.

The notion that effortful encoding is necessary to represent even centrally attended objects gains support from recent work showing that 10-month-old preverbal infants seem unable to detect changes to a moving object that occur during an occlusion event (Xu & Carey, 1996). In these studies, an object (e.g., a toy duck) passes behind an occluding screen and another object (e.g., a toy truck) appears on the other side. Patterns of dishabituation and preferential looking show that 10-month-old infants do not use object properties to determine how many objects are in the display. Interestingly, 12-month-old infants who understand the objects' labels successfully detect the change ' suggesting that verbal or effortful processing is necessary for change detection even when the object is the center of interest (Simons, 1996; Xu & Carey, 1996). The need for effortful, attentional processing to detect object changes suggests that we do not rely on representations of objects or their properties to perceive the continuity of dynamic events. However, both infants and adults can use spatial information to track objects over time, across discrete views, and across occlusion events (Simons, 1996; Spelke, Kestenbaum, Simons, & Wein, 1995; Xu & Carey, 1996).

On the basis of these findings, we reasoned that if tracking the properties of attended objects requires explicit coding, then it should be possible to induce change-detection errors in attended objects if observers did not feel the need to explicitly label their features. Change-detection failures would therefore reflect between-view representations which contain little property information, both for scenes as a whole and for attended objects. To test for change-detection failures, we created a series of short videos in which objects changed either at unpredictable locations or in the center of attention. Experiment 1 was designed to assess the degree to which change-detection failures would occur in edited motion pictures when the objects changed were not necessarily in the center of attention. In Experiments 2A and 2B, change detection for attended objects was examined. In these experiments, observers viewed videos depicting simple events in which a single actor changed into a different person across an edit. In both cases, if object properties were not automatically retained across views, then a significant proportion of changes should go unnoticed.

EXPERIMENT 1

Although researchers have only recently begun to examine between-view representation of natural scenes, film makers have long known that viewers do not store all of the visual details of each view of a scene (Kuleshov, 1920/1987). Fiction films typically consist of events that are presented piecemeal through a series of partial views. These views are often filmed individually, and usually in a different order than depicted in the finished movie. This reordering process can lead to substantial mismatches in object properties, body positions, and clothing across views, yet these inconsistencies (referred to as "continuity errors") are rarely noticed by audiences (Hochberg, 1986; Kuleshov, 1920/1987).

To demonstrate empirically the undetectability of such continuity errors, we created a short edited video of a conversation between two actors. Each cut from one shot to the next resulted in at least one intentional continuity error (see Figure 1). For example, in one case, the red plates on the table changed to white across a cut (Figure 1C-1D). We tested subjects' ability to detect such errors, both when they were unaware that they might occur and when they had been explicitly instructed to look for them.

Method

Ten Cornell University undergraduates viewed the test video in exchange for candy. These subjects viewed a short color VHS video depicting a conversation between two actors. The video included sound and was shown on a 19-in. television with a viewing distance of approximately 100 cm. It initially showed a side view of both actors (Figure I A) and then cut to close up shots of each in turn as they spoke (1B-1D). Across each cut, we made at least one continuity error, for a total of nine intentional errors in the film. For example, in one shot an actor is wearing a large colorful scarf and in the next it has instantaneously disappeared (1A-1B; see also the Appendix). Other changes shown in the figure include a switch from red to white plates (I C-1D) and a change in hand position (I C-1 D). No blanks were inserted between cuts, so the changes occurred within the 60-Hzrefresh rate of a conventional TV (30 two-field frames per second). The subjects who viewed the film were instructed to "pay close attention" but were not forewarned of the changes. After they had viewed the film, they were handed a form that asked, "Did you notice any unusual differences from one shot to the next where objects, body positions, or clothing suddenly changed." They responded by circling "yes" or "no" on the response sheet. The subjects who circled "yes" were then asked to "describe any changes you noticed."

After writing their responses, the subjects were told, "in the film you just viewed, changes occurred every time the camera angle changed. Now I'll show you the film again. This time, try to spot the changes." The question on the first part of the response form cued them to the sorts of changes that might occur (i.e., "objects, body positions, or clothing"). After viewing the film a second time, the subjects were again asked to describe any changes they had noticed.

Results and Discussion

Of the 10 subjects who viewed this film, only I claimed to notice any of the 9 changes during the first viewing. Thus, 89 of 90 total changes went unnoticed. The single detected change was vaguely described as a difference in "the way the people were sitting." Even during the second viewing, when subjects were explicitly told to look for changes, they noticed an average of only 2 of the 9 changes. The most frequently noticed change was the appearance and disappearance of the scarf, which was detected by 7 of the 10 subjects. These results parallel recent findings of failures to detect changes to scenes across views (Blackmore et al., 1995; Grimes, 1996; O'Regan et al., 1996; Rensink et al., 1996; Tarr & Aginsky, 1996), and suggest that edited motion pictures provide a compelling medium for studying naturalistic scene perception and representation.

Although this experiment suggests that we generally do not integrate sensory information across views, we appar-

Figure 1. Four sample views from the stimulus film in Experiment 1. The figure is also available at http://www.wjh.harvard.edu/~dsimons on the World-Wide Web.

ently can form longer lasting representations through effortful encoding-subjects were able to notice a subset of changes when they intentionally searched for continuity errors. Those attended objects are processed more thoroughly, leading to richer representations. These findings are entirely consistent with the limited-capacity serialsearch model discussed earlier; we cannot fully process all objects and properties in a scene (Neisser, 1967), and we are more likely to detect changes to central objects that garner more attentional resources.

This finding again confirms the notion that attention is necessary for change detection. However, it does not directly assess the possibility that attention may not be sufficient for change detection. The next experiment tested the hypothesis that simply focusing attention on an object might not automatically lead to accurate change detection.

EXPERIMENT 2A

If continuous perception is based on representations of an object's spatial position and motion but not its static properties, then even changes to attended objects may go undetected when spatiotemporal information does not signal a change. We tested this hypothesis by creating new motion pictures which focused attention on the very object that changed while maintaining a consistent direction of object motion across cuts. These films portrayed a simple action performed by a "single" actor. But rather than changing small objects which might or might not be attended, we changed the actors themselves. For example, in one film, an actor sitting at a desk hears a telephone ring, gets up, and moves toward the door. The camera then cuts to a view of the hallway where a different actor walks to the telephone and answers it (Figure 2). In Experiment 2A, each observer viewed one of these films with no prior instructions and then wrote a brief description of the film. Experiment 2B was a test of the discriminability of the two actors in which subjects were forewarned that changes would be occurring. This was necessary to ensure that the differences between the actors in the videos were large enough to be plainly visible.

Method

A total of 40 Cornell University undergraduates participated as subjects in Experiment 2A in exchange for candy or course credit. For this experiment, we created a series of eight color VHS videos following the conventional editing practice of maintaining direction of

Figure 2. Sample frames from an actor-change film used in Experiment 2. The figure is also available at http://www.wjh.harvard.edu/~dsimons on the World-Wide Web.

body motion from one shot to the next. The videos were silent and were shown on 19- and 13-in. televisions with a viewing distance of approximately 60-100 cm. Each film showed one of two events: (1) an actor sitting at a desk hears a phone ring in the hallway and gets up to answer it, or (2) an actor enters a room and sits down in a chair. Videos of Event I included two shots (see Figure 2). Videos of Event 2 included three shots. In the first shot, an actor walks through a previously closed door and passes the camera. The second shot shows a wide angle view of the actor walking toward a chair at the front of a classroom while sidestepping other chairs. The final shot is a closeup of the actor sitting in the chair. In each video, the actor initially in the scene changes to a different actor across a cut. (In videos of Event 2, the actor changes between the second and third shots.) Two different pairs of actors were filmed in each event, and separate films were made, with each actor of a pair changing into the other. Each pair of actors was matched for gender, race, hair color, and glasses, and wore globally similar but not identical clothing.

Each of the eight films was viewed by 5 different subjects who were simply asked if they would be willing to watch the video before participating in another study in the same laboratory. They were given no other instructions. The subjects completed the experiment individually. After viewing one of the videos, a subject was given a response sheet that asked him/her to "please write a brief description of the video you saw." The experimenter then took the response sheet from the subject. If a written response did not mention the change from one actor to another, the experimenter directly asked the subject if he or she had noticed the person change. For example, subjects viewing event A were asked "Did you notice that the person who was sitting at the desk was a different person than the one who answered the phone?" Responses to the follow-up question were marked on the response sheet by the experimenter.

Results and Discussion

Viewers were surprisingly oblivious to the substitution of one actor for another, even though the person portrayed by the actors was the central object in the film. Only 33% of the 40 subjects reported that one actor had changed into another. The subjects who failed to notice the change had clearly attended to the films; they produced rich descriptions of clothing, the environment, and the motions of the actors. For example, one subject wrote "Man in light blue shirt & tee shirt was sitting at cluttered desk, turned toward camera & walked toward it. [He] walked outside into hall and picked up the telephone on the wall." In this particular video, the first actor was wearing an unbuttoned, blue, long-sleeve shirt with a T-shirt underneath. The second actor was wearing a gray long-sleeve shirt which was fully buttoned and had no T-shirt underneath. Of those who did not mention the change in their descriptions, only 2 claimed to have noticed it when asked.

EXPERIMENT 2B

Although pretesting had shown that people familiar with the actors in the films used in Experiment 2 immediately noticed the change from one actor to another, we nonetheless empirically controlled for the possibility that the pairs of people in the films were simply too similar to allow detection of the changes. In this experiment, a different set of subjects viewed all of the actor-change videos from Experiment 2A intermixed with a set of videos with no actor changes. The subjects were asked to indicate which films contained a change.

Method

A total of 10 Cornell undergraduates participated as subjects in this study in exchange for candy or course credit. The materials and procedure used in this experiment were identical to those of Experiment 2A with the following exceptions: (I) Eight new videos containing no actor substitutions were added to the 8 actor-change videos from Experiment 2A, for a total of 16 videos. (2) Each subject viewed all 16 films shown in one of two orders (the first was randomly generated and the second was a reversal of the first). (3) The subjects were forewarned that changes would take place in some of the films and were asked to circle "change" or "no change" on a worksheet for each film. They were also instructed to be especially careful, because detecting the change might be more difficult than it sounded. (4) They were informed of the nature of events shown in films of Events I and 2 (see Experiment 2A) using crudely drawn schematic illustrations of each shot in the two videos (i.e., a roughly drawn sketch similar to the panels shown in Figure 2 with stick figures depicting the actors and arrows indicating the direction of motion), and in the case of Event 2 were told that the changes would occur between the second and third shots. Other than the schematic illustration of the cuts, all instructions were given verbally.

Results

Unlike the subjects in Experiment 2A, subjects in this experiment had little difficulty in differentiating the actors. On the average, they made fewer than I mistake each. The inability to notice changes to the central object in the scene cannot be attributed solely to the physical similarity of the objects before and after the change.

GENERAL DISCUSSION

Taken as a whole, these experiments demonstrate that object properties are not automatically used to integrate different views of a scene. Even though we can clearly discriminate individual objects, we apparently do not encode, represent, and use property information to track them over time (Simons, 1996; Xu & Carey, 1996). Our short motion pictures show that even dramatic changes to the very object that is the center of attention often go unnoticed, especially when spatiotemporal information suggests continuity. Although previous research had demonstrated that attention was necessary for the detection of changes and that changes to central objects were detected more readily (O'Regan et aL, 1996; Rensink et al., 1996), interpretations of these results tended to suggest that attention was sufficient for change detection. Along with recent infant work, our studies show that even when we attend to an object, we may not form a rich representation that can be preserved from one view to the next unless the object's properties are intentionally coded. Thus, attending to an object is necessary, but not sufficient, for change detection.

One potentially problematic aspect of our methodology is that, in a sense, we rely on a memory test to reveal on-line detection of a beween-view inconsistency. Given that the observers did not expect the change, our results may reflect the poverty of incidental memory or the difficulty of free recall rather than an inability to detect changes. Although our primary purpose was not to test the ability to *remember vi*sual details of scenes, this research suggests a variety of interesting issues for understanding the relationship between different varieties of memory and change-detection failures. Perhaps a more sensitive test might reveal residual memory for the prechange actor. One possibility is that observers may be able to discriminate the prechange actor from other actors not in the scene when given a recognition or priming task. Although such tasks might reveal the existence of a representation of the prechange actor, such representations probably do not underlie the continuity of our perceptual experience. Presumably, any mechanism underlying a sense of continuity should be tuned toward detecting violations in that continuity. These violations should attract further attention, evaluation, and description, both on the response form and afterwards when observers are directly asked if they had seen the change. Thus, the incidental nature of our task is unlikely to obscure a mechanism that is responsible for tracking objects and detecting violations of expected continuity. However, the present studies do not directly assess the possible existence of other, less accessible representations that might facilitate the processing of repeatedly viewed objects.

Although our findings show that observers often miss surprisingly large changes to central objects, they do sometimes successfully detect changes. One explanation for successful change detection in our task is that observers sometimes intentionally encode (e.g., by verbal labeling) a critical object property just before it changes. The actor changes used in our study did not vary several important categorical properties of the person (e.g., race, sex, age). Perhaps changes to such features would consistently register across views. If so, comparing such successes with the failures we have observed could provide useful insights into the nature of person concepts. These categorical properties could be registered automatically or they might require effortful encoding, but at least for the range of properties manipulated in these studies, change detection seems to require effort and not to be based on automatic domain-general perceptual routines which fully represent the attributes of all attended objects. The visual properties of objects, even attended objects, are not automatically used to integrate different views of a scene.

REFERENCES

BLACKMORE, S. J., BRELSTAFF, G., NELSON, K., & TROSCIANKO, T. (1995). Is the richness of our visual world an illusion? Transsaccadic memory for complex scenes. Perception, 24, 1075-1081.

BRIDGEMAN, B., & MAYER, M. (1983). Failure to integrate visual information from successive fixations. Bulletin of the Psychonomic Society, 21, 285-286.

GRIMES, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (Ed.), *Perception (pp.* 89-110). New York: Oxford University Press.

HOCHBERG, J. (1986). Representation of motion and space in video and cinematic displays. In K. R. Boff, R. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. 1. Sensory processing and perception (pp.* 22-1 to 22-64). New York: Wiley.

INTRAUB, H. (198 1). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception & Performance, 7,* 604-610.

IRWIN, D. E., BROWN, J. S., & SUN, J. S. (1988). Visual masking and visual integration across saccadic eye movements. *Journal of Experimental Psychology: General, It 7,* 276-287.

IRWIN, D. E., YANTIS, S., & JONIDES, J. (1983). Evidence against visual integration across saccadic eye movements. *Perception & Psychophysics, 34,* 49-57.

JONIDES, J., IRWIN, D. E., & YANTIS, S. (1983). Failure to integrate information from successive fixations. *Science, 222,* 188.

KULESHov, L. (1987). *Lev Kuleshov: fifty years infilms. Moscow:* Raduga. (Original work published 1920)

MCCONKIE, G. W., & CURRIE, C. B. (1996). Visual stability across saccades while viewing complex pictures. *Journal of Experimental Psychology: Human Perception & Perlbrmance, 22,* 563-5 8 1.

GRIMES, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (Ed.), *Perception (pp.* 89-110). New York: Oxford University Press.

HOCHBERG, J. (1986). Representation of motion and space in video and cinematic displays. In K. R. Boff, R. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. 1. Sensory processing and perception (pp.* 22-1 to 22-64). New York: Wiley.

INTRAUB, H. (198 1). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception & Performance, 7,* 604-610.

IRWIN, D. E., BROWN, J. S., & SUN, J. S. (1988). Visual masking and visual integration across saccadic eye movements. *Journal of Experimental Psychology: General, It 7,* 276-287.

IRWIN, D. E., YANTIS, S., & JONIDES, J. (1983). Evidence against visual integration across saccadic eye movements. *Perception & Psychophysics, 34,* 49-57.

JONIDES, J., IRWIN, D. E., & YANTIS, S. (1983). Failure to integrate information from successive fixations. *Science, 222,* 188.

KULESHov, L. (1987). *Lev Kuleshov: fifty years infilms. Moscow:* Raduga. (Original work published 1920)

MCCONKIE, G. W., & CURRIE, C. B. (1996). Visual stability across saccades while viewing complex pictures. *Journal of Experimental Psychology: Human Perception & Perlbrmance, 22,* 563-5 8 1.

MCCONKIE, G. W., & ZOLA, D. (1979). Is visual information integrated

across successive fixations in reading? *Perception & Psychophysics, 25, 221-224.*
NEISSER, U. *(1967). Cognitive psychology.* New York: AppletonCentury-Crofts.
O'REGAN, J. K., RENSINK, R. A., & CLARK, J. J. *(1996).* "Mud splashes" render picture changes invisible. *Investigative Ophthalmology & Visual Science, 37, S213.*
PASHLER, H. *(1988).* Familiarity and visual change detection. *Perception & PsYchophysics, 44, 369-378.*
POTTER, M. C. *(1976).* Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Language & Memory, 2, 509-522.*
RENSINK, R. A., UREGAN, *J.* K., & CLARK, *J. J. (1996). To* see or not to see: The need for attention to perceive changes in scenes. *Investigative Ophthalmology & Visual Science, 37, S213.*
SIMONs, D. *J. (1996).* In sight, out of mind: When object representations fail. *Psychological Science, 7, 301-305.*
SPELKE, E. S., KESTENBAUm, R., SIMONs, D. *J., & WEIN,* D. *(11995).* Spatiotemporal continuity, smoothness of motion and object identity in infancy. *British Journal of Developmental Psychology, 13, 113-142.*
TARR, M., & AGINSKY, V. *(1996, July). From objects to scenes: Speculations on similarities and differences.* Paper presented at the Scene Recognition Workshop, Max- Planck- Institut fur Biologische Kybernetik, TCibingen, Germany.
Xu, F., & CAREY, *S. (1996).* Infants' metaphysics: The case of numerical identity. *Cognitive Psychology, 30, 111-153.*

APPENDIX
Description of the Video in Experiment 1

The film used the following six shots to depict a conversation between two actors:

Shot 1: A side view of the two actors (Figure 1A). The actors are sitting at a table with bright red plates, cups, and soft drink bottles on it. The plate in front of actor A (Figure 1B) has food on it. Actor As right hand is resting on the table and she is wearing a colorful scarf. Actor Bs hands are crossed with her elbows resting on the table. The table, both actors, and all of the objects are visible.

Shot 2: A medium shot depicting actor A (Figure 1B) from over the shoulder of actor B. The back of actor Bs head and part of her right shoulder are visible. The plates, food, and utensils are visible as well. Actor A is no longer wearing a scarf.

Shot 3: A medium shot of actor B (Figure 1C from over the shoulder of actor A. The back of actor B's head and part of her shoulder are visible. Her right arm is visible as well. Again, the plates, food, and utensils are visible. Actor A's scarf has returned and actor Bs hand is now on her chin.

Shot 4: A full side view of both actors (as in Shot 1). The plates on the table are now white (rather than red) and actor Bs arms are again crossed with elbows resting on the table.

Shot 5: A medium shot of actor A. The plates are red again, and actor As right hand now rests on her lap rather than on the table.

Shot 6: A medium shot of actor B. The food previously on actor A's plate is now on actor Bs plate, and actor B's right hand again rests on the table.

(Manuscript received September *5, 1996;*
revision accepted for publication April *21, 1997.)*