Review

The representation of visual scenes

Helene Intraub

The visual world exists all around us, yet this information must be gleaned through a succession of eye fixations in which high visual acuity is limited to the small foveal region of each retina. In spite of these physiological constraints, we experience a richly detailed and continuous visual world. Research on transsaccadic memory, perception, picture memory and imagination of scenes will be reviewed. Converging evidence suggests that the representation of visual scenes is much more schematic and abstract than our immediate experience would indicate. The visual system may have evolved to maximize comprehension of discrete views at the expense of representing unnecessary detail, but through the action of attention it allows the viewer to access detail when the need arises. This capability helps to maintain the 'illusion' of seeing a rich and detailed visual world at every glance.

Visual information exists all around us, but physiological constraints prevent us from seeing it all at once. Input is relatively piecemeal as eye movements (called saccades) bring a different region of the world into view as rapidly as three times per second[1]. Furthermore, even when the eye is fixating an area, high resolution is limited to the relatively small area that falls on the fovea. Yet, in spite of these constraints on visual input, viewers claim to experience a clearly visible, detailed and continuous visual world. A classic question in perception has been how this limited input is capable of yielding such a rich visual experience.

Eye movements and scene perception

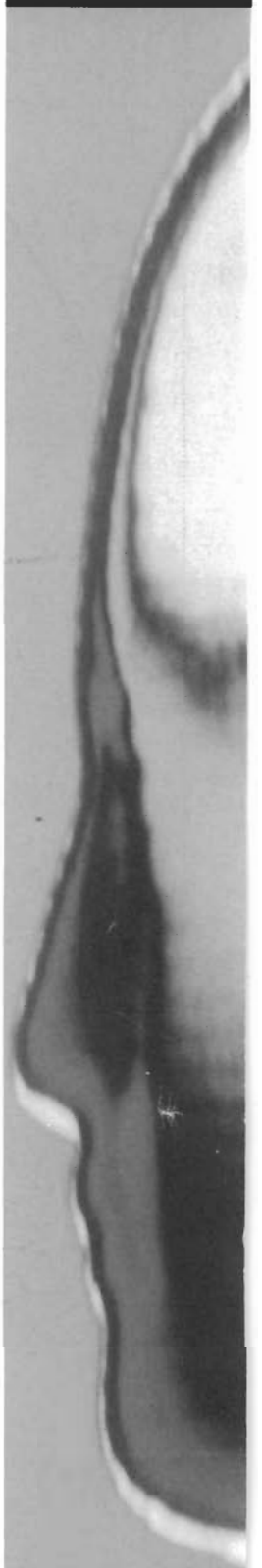
One traditional explanation of our visual experience is that sensory information from each fixation is integrated in a high capacity memory buffer. This information is essentially 'knitted together' into a detailed spatiotopic representation of the environment that is maintained across saccades[2–4]. As we will see, recent research on transsaccadic memory has not supported this theoretical perspective[5–9]. Instead, it supports a somewhat counterintuitive view that the representation of visual scenes is, in large part, abstract and schematic. As the eye briefly shifts from one location to another, the visual world is represented in the form of a schematic map of the scene's layout and major landmarks. Attended areas may be represented in detail, but areas outside the locus of attention are not.

Some of the most compelling evidence for an abstract transsaccadic representation has been obtained in research on reading[7–10]. In one study, readers viewed sentences in which the words were presented in AlTeRnAtInG CaSe on a computer screen[10]. Eye movements were monitored, and during some saccades, the case of every letter was changed. Surprisingly, this did not disrupt the viewers' eye-movement patterns or ability to read. In fact, none of the viewers noticed anything unusual. Clearly information important to normal reading was retained and integrated across fixations, but it apparently did not include the actual visual characteristics of the letters. Similar results have been obtained using visual tasks other than reading[5,11,36].

Recently, photographs of scenes were presented using the same paradigm[12]. In this case, during some saccades, the scene was shifted horizontally or vertically by 0.3°, 0.6° or 1.2° or was expanded or contracted by either 10% or 20%. Eye movements were monitored, and subjects were required to indicate if they noticed a change. Small changes frequently went unnoticed, suggesting that the observers were not using an independent metric of space to piece together successive views. Results suggested that detection of change relied more on local information in the region of the eye's landing position than on a detailed global representation of previously fixated areas.

Though somewhat surprising, given the observer's experience of a detailed visual world, these results fit well with the proposition that not only the transsaccadic representation, but perception itself, is not uniformly detailed and concrete[13–17]. Julian Hochberg has proposed that even during the time that an object is in view, all of its parts are not equally represented in our perceptual system. Specific local and global features of the visual world are fitted into a schematic map of the spatial layout. This schema incorporates detailed attended information, vague global information (e.g. gross size and shape) and expectations about information just outside the field of view. An important aspect of this theory is that the representation is not a fixed and detailed model of the external environment, but a malleable one that continually changes as the viewer shifts attention.

He has supported this argument with many ingenious demonstrations, including the seemingly simple one shown



H. Intraub is at the Department of Psychology, University of Delaware, Newark, DE 19716, USA.

tel: +1 302 831 8012
fax: +1 302 831 3645
e-mail: intraub@udel.edu

 1364-6613/97/$17.00 PII: S1364-6613(97)01067-X

in Fig. 1 (see Refs 14,17). Figure 1A shows a Necker cube, the well-known ambiguous figure that seems to shift perspective as the viewer gazes at it. The usual explanation focuses on the ambiguity of the global stimulus, that allows the mind to shift between two equally plausible alternatives. In Fig. 1B, is a nonambiguous version of a Necker cube, in which the interposition at point 1 dictates a single orientation. According to a wholistic model of perception, such a stimulus should not shift because when the whole object is seen, the interposition will be seen as well. However, contrary to this prediction, if the viewer gazes at point 2, the cube will readily shift. In doing so, as the viewer can attest, point 1 is still seen, it just doesn't seem to exert any influence on the orientation. Hochberg argues that, 'unlike objects themselves, our perception of objects are not everywhere dense...'[17]. That is, even while looking at the cube, it is not represented as a whole – an attended location is represented in a more concrete manner than a location that is not. This description of differences in density, is also captured in recent descriptions of transsaccadic memory and perception, that use the concept of a master map of locations and object files[18–21].

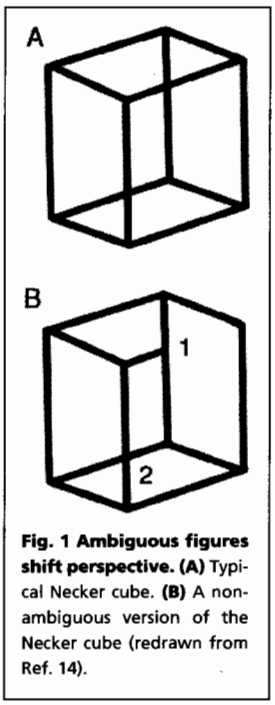


Fig. 1 Ambiguous figures shift perspective. (A) Typical Necker cube. **(B)** A nonambiguous version of the Necker cube (redrawn from Ref. 14).

Applying this concept to dynamically changing scenes in motion pictures, Hochberg argues that it explains why viewers are poor at detecting occasional continuity errors in edited films[22]. Recent research has empirically tested and supported this observation. In one experiment, a continuity error was deliberately created in a videotape of two actors at a table, talking. When the camera panned away from the table to focus on one actor for 4 seconds, a central item on the table (a large soda bottle that had been conspicuously used) was replaced with a cardboard box. When the camera returned to the original view, remaining for 30 seconds, none of the viewers noticed that the bottle had 'become' a box[23].

This inability to detect change was observed even when viewers were explicitly directed to do so, and an object repeatedly changed while the viewer watched a sequence of pictures[24]. The changes were large and often dramatic, covering about 20 square degrees of visual angle, and including the deletion or relocation of an object, or a change in its color. Picture duration was 240 ms with an 80 ms blank interstimulus interval. The initial version was presented twice followed by two repetitions of the changed version, and this alternation pattern continued for as long as the subject needed. Subjects were remarkably poor at detecting the change – in the most difficult cases requiring over 80 alternations (more than 50 seconds) to do so. However, provision of a verbal cue indicating which object was likely to change led to detection within about five alternations or fewer. The changes were easy to see, as long as the viewer was directly attending to the critical location. Although the subjective experience is one of seeing the entire picture, performance suggests instead that at any given moment perception is not 'everywhere dense'.

Although it may seem odd that the visual system would yield a representation that is not 'photograph-like', upon reflection it becomes clear that such a system would be very economical. Viewers would not need to retain a detailed visual representation of the world from moment to moment, because whenever they needed to discern the details of a particular region, they could simply shift their eyes and look at it[6]. The ability to readily do this would support the viewer's impression that the entire visual world is clearly visible at all times. The abstract nature of the representation would also allow for a seamless integration of information that is currently in the visual field with expectations about neighboring information that the next eye fixation is likely to bring into view. With this in mind, we will now consider the implications this has for the long-term representation of scenes, and for imagination.

Remembering and imagining scenes

Does the long-term representation of a scene have a schematic nature similar to what has been proposed for its short-term transsaccadic counterpart, or do these abstract representations in some way summate to create a wholistic memory? If the former is true, then there should be instances in which the long-term representation of a scene will reflect schematic expectations that were never actually viewed. Such cases have been reported. For example, in one study, 24 views of the same city scene were presented: six views (that together yielded a panoramic view) from each of four corners of an intersection[25]. Subjects sorted the 24 pictures, indicating at which of the four locations (shown on a map) they believed the camera had been placed. They received feedback and sorted until reaching a criterion of two correct runs. Later, they were shown new pictures. Each differed by 30 degrees (laterally) from one of the original views. Although they had not seen these particular views of the objects and landmarks before, they were better than chance at reporting the location of the camera. However, in a yes–no recognition test, they were unable to discriminate old from new views. Viewers apparently had retained a schematic representation of spatial layout that had been abstracted from the views in the original sorting task. This representation apparently contained information about the projective sizes, shapes and perspective of numerous landmarks as they would appear from viewpoints that had not been experienced.

Do these remembered expectations about a scene's layout require the scrutiny of numerous views over relatively long periods of time to develop? Or as suggested by Hochberg's formulation, might a single view of a scene be enough to elicit expectations about areas that are not currently in view? 'Boundary extension', a spatial memory distortion for scenes, suggests that even a single view will evoke schematic expectations[26–30], and that it does so quite rapidly.

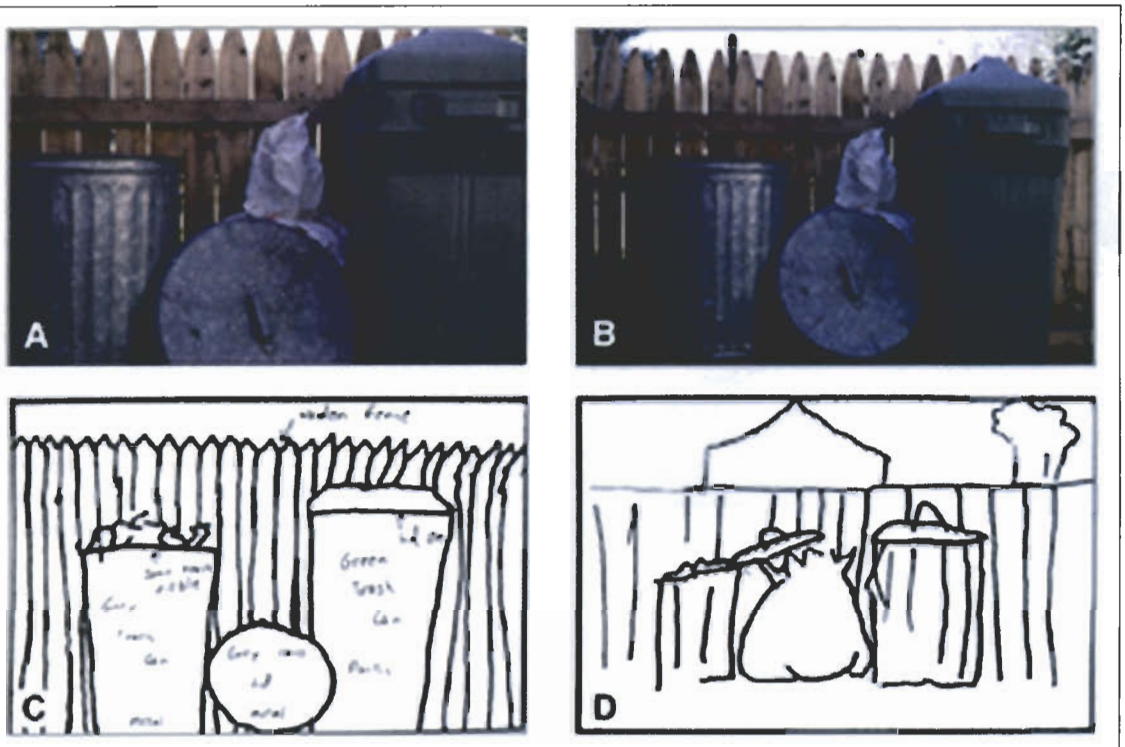


Fig. 2 Examples of boundary extension. When drawing the close-up view **(A)** from memory, the subject's drawing **(C)** contained extended boundaries. Another subject, shown a more wide-angle view of the same scene **(B)** also drew extended boundaries **(D)**. Note: It is important to attend to the edges of a drawing and its associated photograph to see the extent of the distortion. Adapted from Ref. 26.

Boundary extension is a visuospatial memory distortion btained when observers are asked to memorize single views f unrelated scenes. Following presentation, they tend to re-ıember having seen a greater expanse of the scene than was hown in the photograph. Figure 2 provides an example. Echoing Hochberg's claims about schematic expectations, ubjects' memory reflects not only what was physically resent in the picture, but information that was understood o exist just outside the picture's boundaries. The phenom-ıon is evident, not only in drawings but in viewer's re-ponses to recognition test pictures.

When the test picture is the same as the stimulus (a target), ubjects tend to reject it as 'old,' reporting instead that it hows a closer view than did the original. When viewing istractor pictures, subjects show an asymmetric response attern: wide-angle distractors are rated as looking more ke the original picture than are close-up distractors, and vide-angle distractors are more frequently mistaken as being ıe original view[26–30]. Consistent with the notion of spatial xpectarion, both recognition and drawing tests show that oundary extension is greatest for close-ups (in which highly redictable informarion surrounding an attended object is ot physically present), and becomes less apparent as picture iew widens (in which case, the expected information is lready shown in the picture)[27,31]. Wide-angle views often ıow no directional distortion of their boundaries.

Boundary extension suggests a seamless integration of ıformation physically presented in the picture and infor-ıation that was inferred. This is consistent with the notion ıat a single, abstract schematic map underlies both types f information. Might this be the same representation that underlies ttanssaccadic memory? If so, then boundary ex-rension should be detectable very early in processing. In most research, however, picture durations were relatively long (e.g. 15 seconds) and retention intervals ranged from 3 minutes to 2 days. However, recent research supports this possibility in that boundary extension was obtained follow-ing stimulus durations as brief as a single eye fixation (e.g. 250 ms), and at rates of change that simulate rapid visual scanning (e.g. three pictures/second)[29]. In one experiment, on each of 72 trials, three complex color scenes were pre-sented on a computer screen for 333 ms each in rapid suc-cession, followed by a 1 s visual noise mask, and a repetition of one of the three pictures. Subjects tended to rate the repetition as looking 'closer-up' than before, indicating that boundary extension occurs very rapidly indeed.

If boundary extension is due to the activation of ex-pectancies evoked by a partial view of a continuous scene, then it should not occur for stimuli that do not present a partial view[30,32]. To test this possible constraint, memory for spatial expanse was tested for close-up and wide-angle pictures that clearly depicted part of a continuous world (photographs of scenes, and line drawings of the same scenes) and pictures that did not (line drawings of the main object from each scene on a blank background) (see Fig. 3).

Consistent with the hypothesis, photograph-scenes and outline-scenes yielded the boundary extension pattern de-scribed earlier. Close-ups yielded boundary extension, wide-angle views yielded no directional distortion, and responses ro distractors yielded the expected response asymmetry. However, pictures of outline-objects did not yield a unidi-rectional bias. Responses to close-up and wide-angle targets

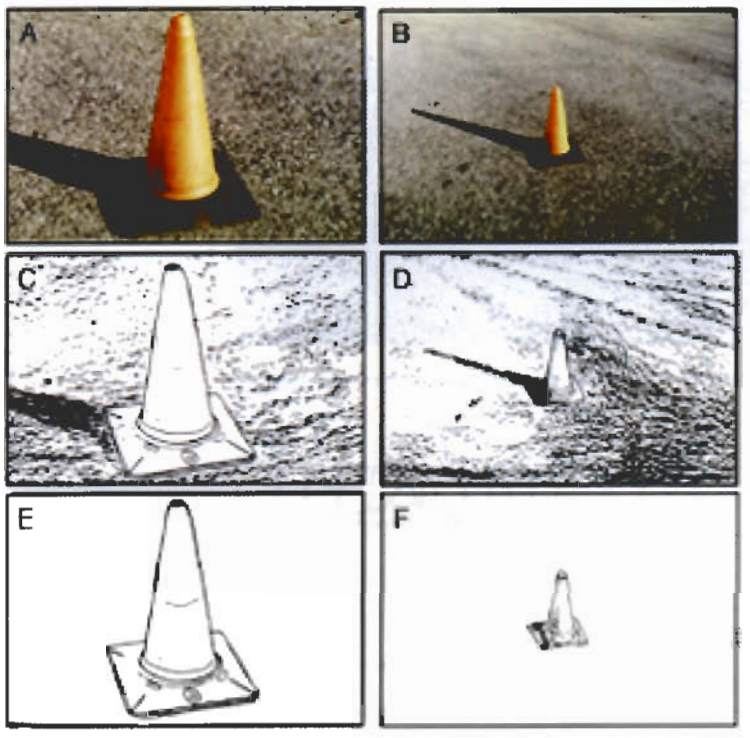


Fig. 3 Close-up and wide-angle views of a traffic cone in a street. (A,B) Photographed scene, **(C,D)** outline-scene, **(E,F)** outline-object. Stimuli from Ref. 30.

were symmetrical. Large objects (close-ups) were remembered as slightly smaller, and small objects (wide-angle views) were remembered as slightly larger (see Fig. 4). Responses to the distractors were also symmetrical. When the background was blank, memory showed evidence of a 'regression to the mean' in terms of object size, rather than a unidirectional distortion of the picture's boundaries to include 'more of the scene'. This suggests that schema activation requires the expectation of a continuous background that is understood to 'exist' just outside a given view.

In a sense, activation of expectations beyond a picture's boundaries is similar to imagining the spatial layout of the scene from which the picture was taken. This raises the possibility that the same abstract mental schema that may underlie perception and memory for scenes may underlie imagination of scenes as well. Behavioral and neuropsychological evidence has supported the theory that perception and imagination share some of the same mental structures[33–35]. In support of this position, the same pictures of outline-objects did yield the typical pattern of boundary errors associated with scenes, when specific 'imagine-scene' instructions were added to the standard instruction[30].

In this case, while memorizing the size of each outline-object, subjects were read a verbal description of the background from each associated photograph and were asked to imagine it while memorizing the picture. Responses to the close-up and wide-angle targets and distractors changed dramatically, yielding the same unidirectional pattern of errors as had the scenes (see Fig. 4). A control condition, in which subjects were read a description of the object's colors to imagine, showed that this change in memory for spatial expanse was not due to the introduction of an imagination task per se. In the imagine-colors condition, as in the initial standard condition, the unidirectional bias indicative of boundary extension was eliminated (see Fig. 4).

These experiments demonstrate that the patterns of errors in memory for layout and object size were not dictated by the physical stimulus, but were determined by whether or not the viewer understood the display to be part of a continuous scene – a context that activates schematic expectations.

Outstanding questions

- What is the best way to characterize the notion of an 'abstract schematic representation' of a scene?
- What type of model would be most useful for expressing the effects of attention on changes in information density across the representation over time?
- Can the concept of 'object files' be used as a means of capturing the complex relations among objects and background elements in a scene?
- What types of tests would be most convincing in establishing whether or not transsaccadic memory, perception, picture memory and imagination of scenes share the same underlying representation, or are just similar processes?
- Do many 'filling-in' processes in visual cognition share the same underlying mechanisms? (e.g. amodal completion, completion across the blind-spot[37], and boundary extension?)

Conclusions

Research on transsaccadic memory, perception, picture memory and imagination yield support for the notion of an abstract schematic representation of visual scenes. However, critical questions remain to be answered before a strong theoretical stance can be taken. Foremost is the need to formally specify what is meant by an 'abstract mental schema'. To this end, research is needed that will focus on: (a) the variability in the 'density' of information across the representation, (b) the best means of modeling the way in which this 'density map' changes as attention shifts to different regions of the representation and (c) the extent to which this model can predict performance during perception, retrieval and imagination of scenes. At this stage, the notion of a common underlying representation for such a wide range of cognitive functions is highly speculative. Does a common representation underlie transsaccadic memory and boundary extension? Perhaps it does, or perhaps the similarity between the two is relatively superficial. To pursue this question, it is necessary to obtain converging evidence regarding the effect of a number of factors on both. For example, if a picture was removed from view just as a fixation to the right was about to be implemented, would there be a rightward bias in boundary memory for the scene? And if such dynamic effects are found in picture memory, will they also be manifest in an imagination task? At present, what the reviewed research does clearly show is that the representation of visual scenes is much more schematic and abstract than our immediate experience suggests. The visual system has evolved in such a way as to maximize comprehension of discrete views at the expense of unnecessary detail, but through the action of attention allows the viewer to access detail when the need arises. Trends in the literature suggest that the visual system may treat all views in the same way, regardless of their source: an eye fixation, a picture, or an act of imagination.

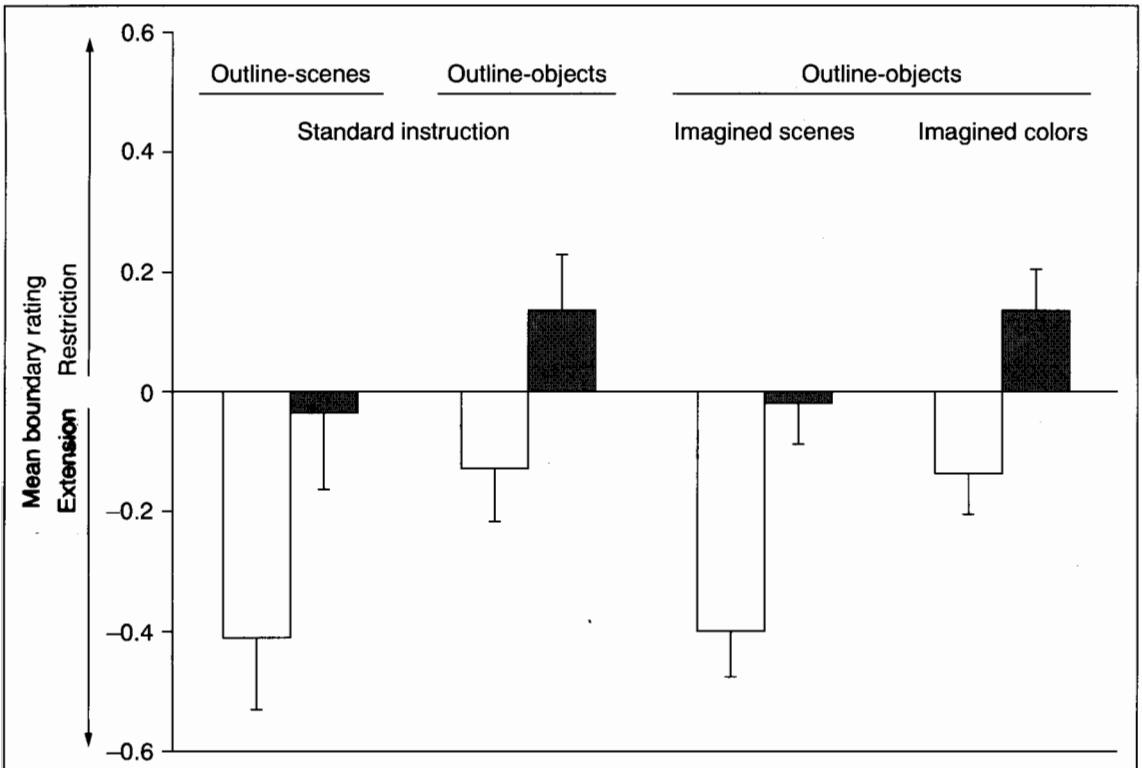


Fig. 4 Error patterns in memory for scenes and objects. Mean boundary ratings and 0.95 confidence intervals for close-up (white bars) and wide-angle (grey bars) target pictures in the outline-scene and outline-object conditions following standard (no-imagery) instructions, and for the outline-object conditions when subjects either imagined scenes or imagined the objects in color. Test pictures were rated on a 5-point scale ranging from –2 (too much of a close-up) to +2 (too much of a wide-angle view). Negative scores that differ from 0 indicate boundary extension, and positive scores that differ from 0 indicate boundary restriction. Bar graphs show results from Experiments 1, 3 and 4, Ref. 30.

Acknowledgements

This review was supported by Grant MH54688 from the NIMH. I thank Carmela V. Gottesman and Barbara Landau for their helpful comments on an earlier draft.

References

1 Yarbus, A. (1967) *Eye Movements and Vision*, Plenum Press
2 Davidson, M.L., Fox, M.J. and Dick, A.O. (1973) Effect of eye movements on backward masking and perceived location *Percept. Psychophys.* 14, 110–116
3 McConkie, G.W. and Rayner, K. (1976) Identifying the span of the effective stimulus in reading: literature review and theories of reading, in *Theoretical Models and Processes of Reading* (Singer, H. and Ruddell, R.B., eds), pp.137–162, International Reading Association, Newark, NJ
4 Breitmeyer, B.G., Kropfl, W. and Julesz, B. (1982) The existence and role of retinotopic and spatiotopic forms of visual persistence *Acta Psychol.* 52, 175–196
5 Irwin, D.E. (1991) Information integration across saccadic eye movements *Cognit. Psychol.* 23, 420–456
6 O'Regan, J.K. (1992) Solving the 'real' mysteries of visual perception: The world as an outside memory *Can. J. Psychol.* 46, 461–488
7 Pollatsek, A. and Rayner, K. (1992) What is integrated across fixations, in *Eye Movements and Visual Cognition: Scene Perception and Reading* (Rayner, K. ed.), pp. 166–191, Springer-Verlag
8 Rayner, K. and Pollatsek, A. (1992) Eye movements and scene perception *Can. J. Psychol.* 46, 342–376
9 Irwin, D.E. (1993) Perceiving an integrated visual world, in *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience* (Meyer, D.E. and Kornblum, S., eds), pp.121–142, MIT Press
10 McConkie, G.W. and Zola, D. (1979) Is visual information integrated across successive fixations in reading? *Percept. Psychophys.* 25, 221–224
11 Bridgeman, B., Hendry, D. and Stark, L. (1975) Failure to detect displacement of the visual world during saccadic eye movements *Vis. Res.* 21, 285–286
12 McConkie, G.W. and Currie, C.B. (1996) Visual stability across saccades while viewing complex pictures *J. Exp. Psychol. Hum. Percept. Perform.* 22, 563–581
13 Rock, I. (1977) In defense of unconscious inference, in *Stability and Constancy in Visual Perception: Mechanisms and Processes* (Epstein, W. ed.), pp. 321–373, Wiley
14 Hochberg, J. (1978) *Perception* (2nd edn), Prentice-Hall
15 Rock, I. (1997) *Indirect Perception*, MIT Press
16 Hochberg, J. (1981) On cognition in perception: Perceptual coupling and unconscious inferences *Cognition* 10, 127–134
17 Hochberg, J. (1982) How big is a stimulus? in *Organization and Representation in Perception* (Beck, J., ed.), pp. 191–217, Erlbaum
18 Treisman, A. (1988) Features and objects: The Fourteenth Bartlett Memorial Lecture *Q. J. Exp. Psychol.* 40A, 201–237
19 Kahneman, D., Treisman, A. and Gibbs, B.J. (1992) The reviewing of object files: Object-specific integration of information *Cognit. Psychol.* 24, 175–219
20 Gordon, R.D. and Irwin, D.E. (1996) What's in an object file: Evidence from priming studies *Percept. Psychophys.* 58, 1260–1277
21 Henderson, J.M. and Anes, M.D. (1994) Role of object-file review and type priming in visual identification within and across eye fixations. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 826–839
22 Hochberg, J. (1986) Representation of motion and space in video and cinematic displays, in *Handbook of Perception and Human Performance (Vol. 1)* (Boff, K.J., Kaufman, L. and Thomas, J.P., eds), pp. 22:1–22:64, Wiley
23 Simons, D.J. (1996) In sight, out of mind: When object representations fail *Psychol. Sci.* 7, 301–305
24 Rensink, R.A., O'Regan, J.K. and Clark, J.J. To see or not to see: The need for attention to perceive changes in scenes *Psychol. Sci.* (in press)
25 Hock, H.S. and Schmelzkopf, K.R. (1980) The abstraction of schematic representations from photographs of real-world scenes *Mem. Cognit.* 8, 543–554
26 Intraub, H. and Richardson, M. (1989) Wide-angle memories of