Level 1

Visual Scene Perception

Helene Intraub, University of Delaware, Newark, Delaware, USA

CONTENTS

When studying a scene, viewers can make as many as three or four eye fixations per second. A single fixation is typically enough to allow comprehension of a scene. What is remembered, however, is not a photographic replica, but a more abstract mental representation that captures the 'gist' and general layout of the scene along with a limited amount of detail.

INTRODUCTION

0632.001 The visual world exists all around us, but physiological constraints prevent us from seeing it all at once. Eye movements (called saccades) shift the position of the eye as quickly as three or four times per second. Vision is suppressed during each saccade, and then resumes during the next eye fixation (the period of time that the eye pauses to receive visual input). It has been postulated that the mental representation of a scene that is maintained across a saccade ('transsaccadic memory') is surprisingly sparse. This information, in conjunction with expectations about upcoming layout, provides a means for integrating successive views into a coherent representation of a scene.

SCENE COMPREHENSION AND ITS INFLUENCE ON OBJECT DETECTION

0632.002 Perception occurs so quickly and automatically that individuals cannot discern the amount of time they need to understand a scene. Are multiple eye fixations necessary for scene perception, or are we able to perceive the meaning of a scene right away, based on the first eye fixation? One way to address this question is to see if viewers can understand unrelated scenes presented in rapid succession at a rate that mimics the rapid pace at which we normally make eye fixations (e.g. three per second). The reason for using unrelated scenes is to allow an assessment of what can be gleaned from a single glimpse without any bias from prior context. Immediately after viewing a rapid sequence like this, viewers participate in a recognition test. The pictures they just viewed are mixed with new pictures and are slowly presented one at a time. The viewers' task is to indicate which pictures they remember seeing before. Following rapid presentation, memory for the previously seen pictures is rather poor. For example, when 16 pictures were presented in this way, moments later, viewers were only able to recognize about 40 to 50 per cent in the memory test. One explanation is that they were only able to perceive about 40 to 50 per cent of the scenes that had flashed by and those were the scenes they remembered. However, contrary to this plausible explanation, many viewers adamantly claimed that they had indeed momentarily perceived most of the scenes but that the onslaught of new scenes interfered with their ability to remember what they had momentarily grasped.

To determine whether viewers were good at mo- 0632.003
mentarily grasping the meaning of a scene under these conditions, it was necessary to design a task that could tap into the early stages of scene processing at a point before forgetting is likely to take place. Subjects were shown the same rapid sequences as in the previous experiments, but in this case a brief description of one of the pictures (e.g. 'a road with cars') was provided in advance. They were instructed to look for a picture matching that description and to immediately press a response key as soon as they saw it. In some experiments the description was fairly specific (as in the example 'a road with cars') and in others it was very broad and general (e.g. 'a type of furniture'). Because the detection task required them to respond immediately (without waiting until the end of the sequence), this minimized the likelihood of forgetting. At speeds as fast as three and four pictures per second, the ability to correctly detect a

scene based on a description was excellent, usually reaching about 90 per cent correct or better. This performance far exceeded the 40 to 50 per cent correct obtained from subjects who took the recognition test. These results indicate that scene perception is very rapid: a single fixation is enough to allow the viewer to identify its meaning.

0632.004 Scenes, however, usually contain multiple objects. There are two ways that the identification process might proceed: (a) first, objects are identified, and then viewers begin to understand what kind of scene they are looking at, or (b) first, the scene's general meaning and layout are perceived holistically, and then identification of specific objects follows. Although a controversial topic, many studies suggest the second alternative. In a series of experiments, outline drawings of scenes were each briefly presented (e.g. 50–150 milliseconds) one at a time. Prior to the presentation of each scene, viewers were provided with the name of an object (e.g. fire hydrant), and immediately after each presentation a dot appeared on the screen at the location previously filled by one of the objects in the scene. Viewers had to say whether or not the object named at the outset (in this example, a fire hydrant) had been present at that location. In some cases the object did fit with the general meaning of the scene (e.g. a fire hydrant in a street scene) and in others it did not (e.g. a fire hydrant in a kitchen scene). If objects are identified before the scene's meaning is grasped, then the ability to detect the object should be unaffected by whether or not it fits with the context of the scene. However, contrary to this possibility, the object's relation to the scene as a whole did affect the speed and accuracy of the response. Other relations of the object to the scene, such as whether its relative size did or did not fit the scene, or its location in the scene was plausible or implausible, had a similar effect on the response. This suggests that we may grasp the meaning and general layout of a scene rapidly enough to affect our ability to detect specific objects in the scene.

0632.005 A similar conclusion has been drawn from other experiments in which eye movements were monitored. Pictures were presented for relatively long intervals that allowed viewers to make many eye fixations. Prior to the picture's onset, viewers fixated the centre of the screen. Frequently, the first saccade brought the eyes to an object that didn't belong, and subjects maintained longer fixation times on that object, as if they were trying to understand it. Again, this suggests that a scene's meaning and layout is understood very rapidly in processing, occurring prior to identification of all its objects. How does this information become integrated with new information obtained from the next eye fixation?

TRANSSACCADIC MEMORY

Initially, many researchers who studied the relation 0632.006
between eye movements and cognition thought that after each eye fixation on a scene, a detailed sensory record of the fixated region was stored in a very short-term memory system called an 'integrative buffer'. Within the buffer, information from a new fixation would be integrated (i.e. knitted together) with the stored information from the previous fixation. By integrating the details of successive views, the system was thought to provide the viewer with a seamless, detailed perception of the world – like piecing together parts of a jigsaw puzzle. However, research on memory for briefly glimpsed scenes, and research on reading and eye movements, has led to a different interpretation about how views are integrated across saccades. The idea is that our perception of a detailed, continuous world is to some degree illusory. Instead it is proposed that 'transsaccadic memory' (memory for information that is maintained across a saccade) includes the scene's meaning, along with the general layout, and only some detail.

Transsaccadic memory allows the visual system 0632.007
to relate the information obtained from one eye fixation to the contents of the next – but not by integrating detailed photograph-like views. It is somewhat surprising to think that our perception of the world is built up from relatively sparse information, but there is a lot of evidence to support this claim. You can get a sense of this yourself if you look at a complex visual scene (e.g. a bookshelf with books of different sizes and colors) and then close your eyes and recount all the details. You will probably find that your visual memory is missing a lot of information. Of course, the claim is best supported by controlled experiments that carefully address the viewer's ability to detect changes.

Various types of change detection experiments 0632.008
have been conducted in which a change is made in a display at exactly the same time that the eyes make a saccade. Using computer technology, a viewer's eyes are monitored, and when a saccade is launched, a change is made before the eyes land again and begin the next fixation. Therefore, the change in the scene takes place while vision is suppressed. When the eyes land, the change has already occurred. In these situations, many kinds of changes go unnoticed both in scenes and in text. For example, in one case, researchers presented

sentences in an AlTeRnAtInG cAsE on a computer screen. Each time the viewer made an eye movement, the case was reversed (capital letters became small and small letters became capitals). This did not disrupt their ability to read, and understand the text. Amazingly, the viewers didn't even notice the changes.

PERCEPTION OF 'GIST' VERSUS DETAILS

0632.009 Although the research described earlier shows that viewers can rapidly understand the general meaning or 'gist' of a scene, they are remarkably poor at noticing changes from one look to the next: a phenomenon called 'change blindness'. One of the best examples of change blindness uses a presentation technique (the 'flicker' technique) that is very similar to the rapid picture presentation technique described earlier. So it is interesting to make a comparison.

0632.010 In preparation for the study, the experimenter selects scenes and then uses computer graphics to edit each one, changing a particular feature (e.g. a lamp disappears or changes color; diagonal stripes on a wall are reversed). The original and the altered version are each presented for a brief duration that mimics a single eye fixation. The two versions keep alternating throughout the sequence with a blank interval in between presentations. This interval is meant to simulate the suppression of vision during a saccade. The viewer's task is to identify the change. Even though they can clearly identify the scene every time it appears, they do not notice the change right away. The most difficult cases required more than 80 alternations (i.e. more than 50 seconds) before even large changes were detected. What is most interesting is that the changes were actually very easy to see if the viewer received a hint about the location, or just fixated the location. This demonstrates that memory for a scene, just a fraction of a second later, does not provide us with a detailed, picture-like representation – but, as shown earlier, it does provide us with enough information to understand what we are looking at.

BOUNDARY EXTENSION

0632.011 How does the visual system integrate the relatively sparse information held in transsaccadic memory from one fixation to the next, thus allowing us to understand the whole scene? A phenomenon called 'boundary extension' provides one answer to this perplexing question. It also shows that although the representation of a scene lacks detail, it has an overabundance of another type of information that may serve to facilitate integration of views. After looking at photographs for as long as 30 seconds each, people tend to make an interesting error. They remember having seen beyond the edges of the picture! They remember seeing information that was *not* in the picture but that was likely to have existed just outside the camera's range of view.

0632.012 This overinclusive memory is so convincing, that when they see the same photographs again (in a recognition test) they reject them as being the same: they claim that the test picture doesn't show as much of the scene as had the 'original' picture they studied earlier. Boundary extension can also be seen when other viewers draw the photographs from memory. The first two pictures in Figure 1 (panels (a) and (b)) show a close-up view of a scene and a viewer's drawing from memory. Notice that the drawing extends the boundaries of the view. Although the trashcans, fence, and lid are all cropped by the photograph's edges, the subject remembered seeing them as whole, and also remembered seeing parts of the fence on the left and right of the scene, some sky above the fence, and more of the scene at the bottom. It is tempting to think of this as just an error, but if you look at a wider-angle photograph of the same scene in panel (c), you will see that the viewer's 'error' actually provides an excellent prediction of what really did exist outside the boundaries of the close-up. The drawing looks more like the wide-angle view than the close-up that the subject had actually studied. This effect appears to be the rule rather than the exception in memory for scenes (particularly close-ups). In one experiment it occurred in 95 per cent of the 133 drawings made by 20 different people. Other research showed that even when people tried hard not to make this mistake, they couldn't prevent it from happening.

0632.013 Boundary extension reveals the remarkable ability of the visual system to predict the continuation of scene layout and may serve to facilitate the integration of views by (a) 'priming' the visual system to see the upcoming layout, and (b) placing the view within a larger context. To determine whether boundary extension would occur rapidly enough to aid integration of views during visual scanning, rapid sequences of pictures were again used in an experiment. Subjects viewed three pictures in rapid succession, and one second later, one picture was repeated and remained on the screen. Viewers rated the picture on a 5-point scale to indicate if

the view was the same, was more wide-angle, or more close up than the view they saw a second earlier. Subjects tended to rate the repeated picture as showing less of the scene than the 'original' picture. The visual system very rapidly extrapolated the picture's layout – so that the viewer remembered having seen more of the scene than they actually did. Recent research has shown that the same anticipatory error occurs in memory for real three-dimensional scenes that are viewed through a window. People remember seeing expected information from just outside the view.

0632.014 Taken together, the results of these studies indicate that transsaccadic memory (the memory between saccades) is not highly detailed; it is more abstract and schematic than most people would expect. The ability to (a) perceive and remember rapidly presented scenes, (b) identify an object more readily when it fits a scene, and (c) rapidly extrapolate layout beyond the edges of a view, all show that viewers rapidly grasp the general meaning and the layout of a scene. Change detection experiments demonstrate that many details are left out of the representation, whereas research on boundary extension shows that highly anticipated parts of the background are frequently added. A valuable aspect of this type of representation is that it minimizes the possibility of 'memory overload' while at the same time facilitating the integration of critical aspects of successive views. It has been argued that we don't really need to have a highly detailed visual memory of our environment because if we want to see a detail, we can rapidly fixate the region in question – an act requiring only a fraction of a second. But to accurately fixate the location we need to understand the scene and possess a good sense of its layout. There is considerable economy in the way we perceive and remember the visual world.

Further Reading

Biederman I (1981) On the semantics of a glance at a scene. In: Kubovy M and Pomerantz JR (eds) *Perceptual Organization*, pp. 213–253. Hillsdale, NJ: Lawrence Erlbaum Associates.

Intraub H (1997) The representation of visual scenes. *Trends in the Cognitive Sciences* **1**: 217–221.

Intraub H (1999) Understanding and remembering briefly glimpsed pictures: implications for visual scanning and memory. In: Coltheart V (ed.) *Fleeting Memories*, pp. 47–70. Cambridge, MA: MIT Press.

Irwin DE (1991) Information integration across saccadic eye movements. *Cognitive Psychology* **23**: 420–456.

McConkie GW and Zola D (1979) Is visual information integrated across successive fixations in reading? *Perception & Psychophysics* **25**: 221–224.

O'Regan JK, Rensink RA and Clark JJ (1999) Change blindness as a result of 'mudsplashes'. *Nature* **398**: 34.

Potter MC (1999) Understanding sentences and scenes: the role of conceptual short-term memory. In: Coltheart V (ed.) *Fleeting Memories*, pp. 13–46 Cambridge, MA: MIT Press.

Rayner K and Pollatsek A (1992) Eye movements and scene perception. *Canadian Journal of Psychology* **46**: 342–376.

Simons DJ and Levin DT (1997) Change blindness. *Trends in Cognitive Science* **1**: 261–267.

Wolfe JM (1999) Inattentional amnesia. In: Coltheart V (ed.) *Fleeting Memories*, pp. 71–94 Cambridge, MA: MIT Press.

Glossary

Eye fixation The period of time that the eye remains focused on a single location for the purpose of extracting information before moving to a new location. Viewers can make as many as three or four eye fixations in a second.

Layout The arrangement of objects with respect to one another and to the background features of a scene.

Recognition test The purpose of the test is to determine whether previously seen items look familiar, by re-presenting those items in the context of new items and asking a person to discriminate one from the other.

Saccade A rapid, ballistic movement of the eye that allows the viewer to shift the point of fixation. The saccade takes a fraction of a second to occur, during which time vision is momentarily suppressed.

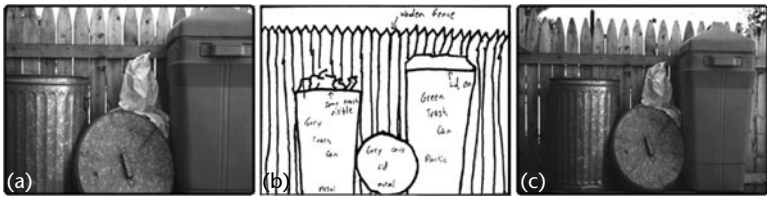
Transsaccadic memory Memory for aspects of a fixated region of a scene (or other visual stimulus) that survives an eye movement (i.e. a saccade).

Visual scene perception The ability to understand a scene and grasp the relations between the objects and surfaces based upon visual input.

Visual suppression When the eye rapidly moves during a saccade, the visual world becomes a blur on the retina. During this time, vision is suppressed so that we don't typically see the blur during visual scanning.

Keywords: (Check)

visual perception; scenes; transsaccadic memory; boundary extension; change blindness



0632f001 **Figure 1.** (a) A photographic close-up of a scene, (b) a viewer's drawing of the close-up from memory (note the extended boundaries), and, (c) a more wide-angle photograph of the same scene. (Based on H Intraub and M Richardson (1989) Wide-angle memories of close-up scenes. *Journal of Experimental Psychology: Learning, Memory and Cognition* **15**: 179–187)

ECSaq632
Queries for Macmillan, ECS paper no. 632

Title: Visual scene perception
Author: Intraub

Section “Perception of ‘gist’ versus details”, end of penultimate sentence: does the altered wording capture the correct sense? i.e. “… if the viewer received a hint about the location, or just fixated the location.”

Figure: please confirm that any necessary permission to reproduce this has been obtained. Also, preferably we need a higher-resolution file (300 dpi or above), do you have access to such a file?