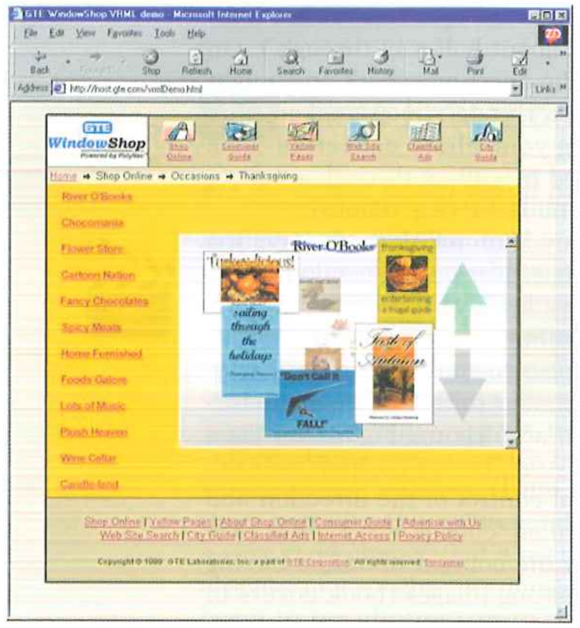

FIGURE 4.40 'Floating RSVP' in which images appear to approach the viewer from a distance. Sensitive arrows allow the speed and direction of 'movement' to be controlled by a user

(Courtesy Kent Wittenburg)

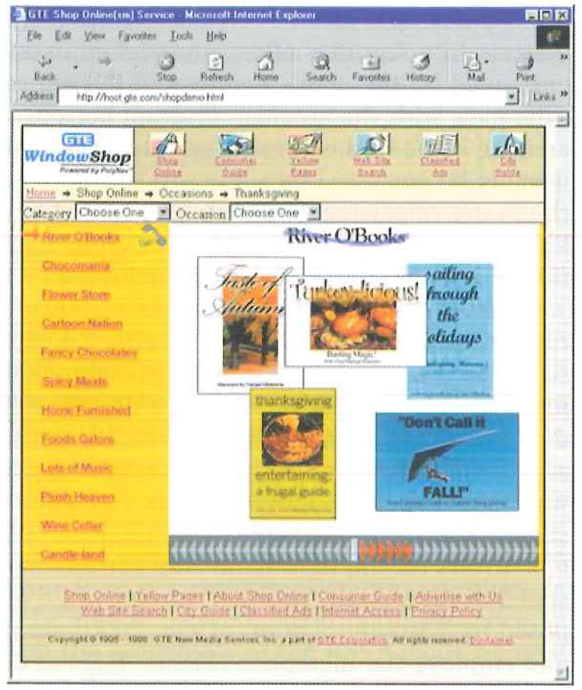

FIGURE 4.41 The contents of an online bookstore are presented in 'collage mode' RSVP, simulating the placing of book covers on a table in sequence. The set of arrows just under the presentation allows control of the speed and direction of presentation

(Courtesy Kent Wittenburg)

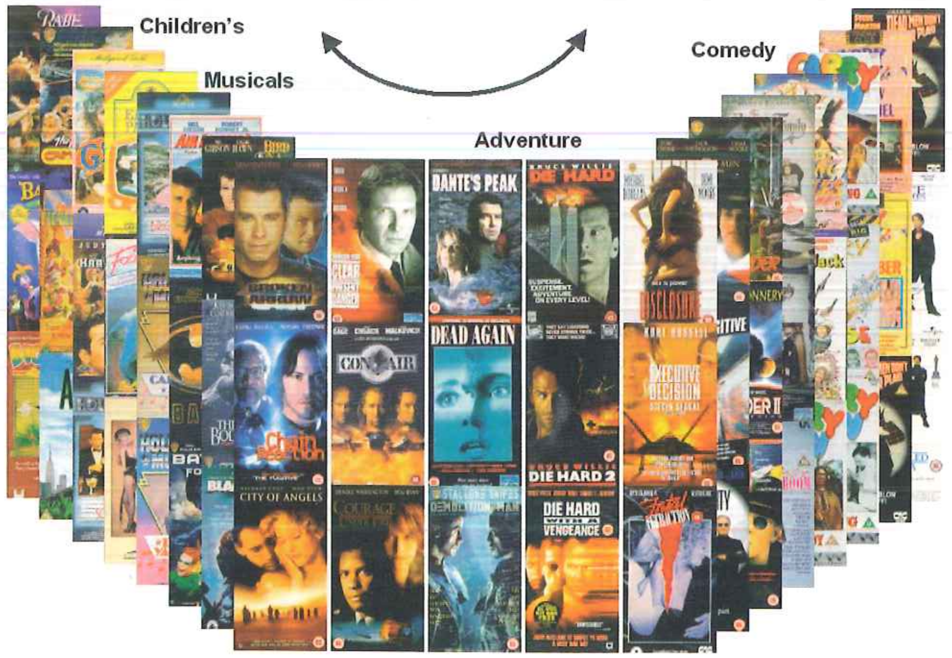

FIGURE 4.42 An interface facilitating the browsing of posters advertising videos. Cursor movement along the stacks causes posters to briefly 'pop out' sideways and the whole bifocal structure can be scrolled to bring a video of interest to the central region, where a mouse click will cause a clip from a video to be played

Briefly glimpsed images

An obvious question is, 'How fast can images be presented while allowing a sought-after (target) image to be identified?' In other words, how rapid is the 'rapid' in RSVP? The faster a collection of photographs can be presented while allowing the single one of interest to be identified, the more attractive is RSVP as a presentation technique for that task. Relevant experimental evidence is available (Potter and Levy, 1969) and is summarized in Figure 4.43. In this experiment, a subject is first shown an image. They are then asked to say whether that image appears in a sequence of images presented, one after the other, in quick succession. It is found that even if the rate of image presentation is as high as ten per second, so that each image is presented for only 100 ms, there is about a 90 per cent chance that the presence or absence of the image in the collection will be correctly reported. The experiment summarized in Figure 4.43, although probably the most pertinent to the application of RSVP, is nevertheless only one source of experimental evidence that has been accumulated over the years and, indeed, continues to be accumulated. For the reader who wishes to 'dig deeper', the book *Fleeting Memories* (Coltheart, 1999b) is highly recommended.

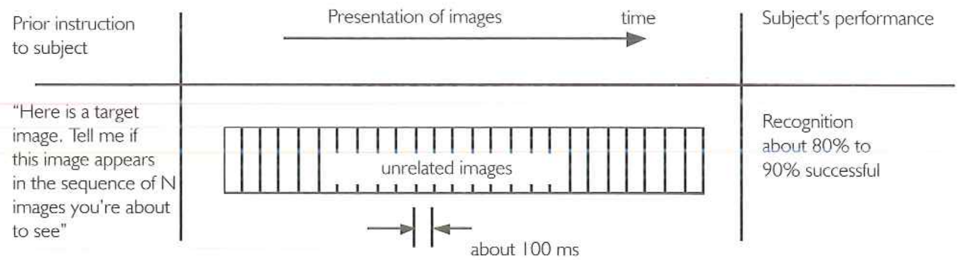

FIGURE 4.43 An experiment to test a subject's ability to recognize the presence or absence of a previously viewed target image within a collection presented sequentially at a rate of around ten per second

4.2.2 *Space and time resources*

It is easy, but dangerous, to argue intuitively about the merits of different presentation modes, some of which have already been encountered. What is needed is some basis for comparison. One such basis is that of *resource*. One could argue that the interaction designer concerned with the presentation of a collection of images to support a task is limited by two resources. One is *display area*, a resource whose exploitation we have just discussed in Section 4.1. Another is *total presentation time*, in the sense that tasks usually have to be performed within a given time limit. We have all had the experience of systems in which the difficulty and frustration of performing a task have been exacerbated by the excessive time required for relevant information to be presented.

We can represent concurrent limits to display area and total presentation time by the 'resource box' shown in Figure 4.44. The interesting question now is, 'Given that a collection of images needs to be presented to a user, how should those images be arranged in the resource box?' There are many known possibilities, some of which we have already met, and others will no doubt be invented, but to present information of value to an interaction designer we select, for illustration, the six presentation modes shown in Figure 4.45. Commensurate with our use of the resource box for comparison, the available display area is assumed to be the same for all modes, as is the total presentation time.

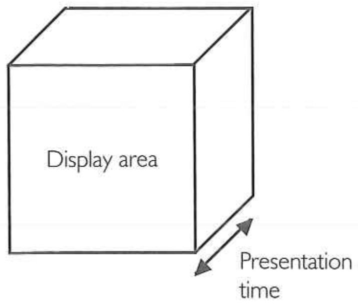

FIGURE 4.44 Representation of limits on display area and total presentation time by a 'resource box'

The modes shown in Figure 4.45 have been arranged into two groups based on a simple criterion: whether any given image appears in a unique position on a display (the 'static' group) or can appear to move by occupying a number of positions in a sequence (the 'moving' group). Of the three static modes we have already encountered slide-show mode (A) and tile mode (C). In the remaining static mode – 'mixed mode' (B) – each image has the possible advantage of being displayed for four times as long as in slide-show mode but the possible disadvantage of occupying one-quarter the display area. It is included in Figure 4.45 because, in early experiments (Fawcett *et al.*, 2004), it was the most preferred static mode and often led to the most efficient identification of the presence or absence of a target image in the presented collection.

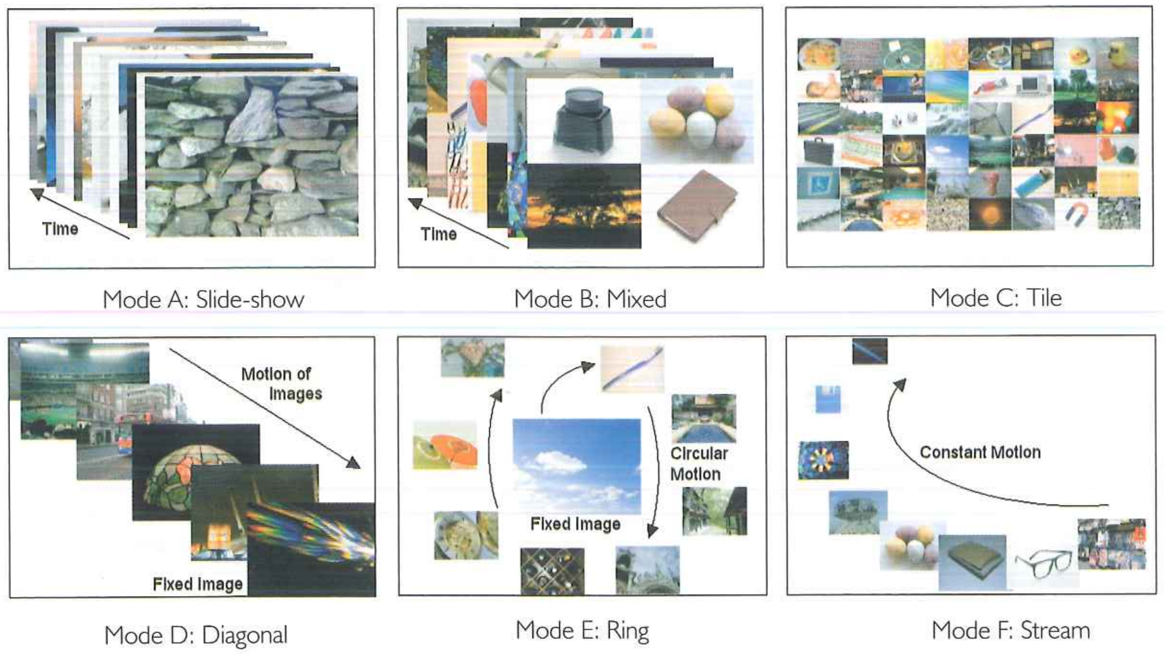

FIGURE 4.45 Three 'static' image presentation modes (A, B, C) and three 'moving' image presentation modes (D, E, F)
Source: *Courtesy of Katy Cooper*

Of the three moving modes shown in Figure 4.45 the first (D), called the diagonal mode – also known as the 'time-tunnel' mode (Wittenburg *et al.*, 2003) – involves movement of each image from top left to bottom right: the interaction designer might choose to let images increase in size as they proceed along the diagonal trajectory and/or have adjacent images overlap to some degree, hoping that the advantage of more images will not be offset by the reduced fraction of each that is visible. With this mode it turns out – as we shall see later – that a crucial design decision concerns the manner in which an image disappears from the display. It could move in a continuous fashion so that less and less of it appears, or it could be 'captured' (Wittenburg *et al.*, 2003) full-size at the lower right-hand corner for about 100 ms before removal from the display. The latter option turns out to be a wise choice, as we shall see. In the ring mode (E), an image first appears at maximum size in the centre of the display, remains there for a short time and then moves in a roughly circular trajectory, exiting top left. In the stream mode (F), images are in constant motion, entering from lower right and exiting towards upper left.

Since the relative benefits of the six modes illustrated in Figure 4.45 appear to be closely associated with the eye-gaze behaviour adopted by a user, we anticipate later discussion by briefly examining this aspect of human visual performance.

Eye-gaze

When we look at a display with the intention of gaining insight, we constantly move our eyes with the result that our eye-gaze travels over the displayed image (Ware, 2004). The eye-gaze of a human being is primarily characterized by **saccades** and **fixations** (Figure 4.46). Saccades are movements of gaze: they are rapid – lasting between about 20 and 100 ms – but also ballistic in the sense that, once initiated, the movement cannot be modified mid-saccade. Fixation refers to a dwelling of gaze at a fixed point: a fixation generally lasts between about 200 and 600 ms. Indeed, we are all familiar with saccades and fixations

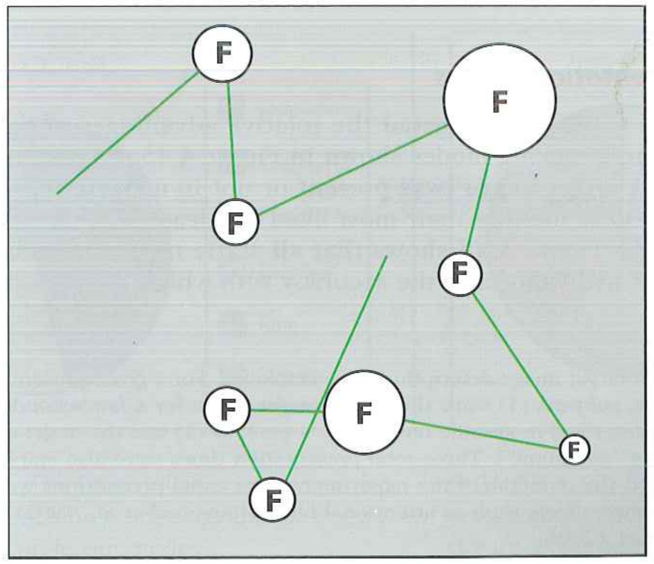

FIGURE 4.46 A simple representation of eye-gaze behaviour. The rapid saccades are shown green, the fixations (F) of varying duration by circles of proportionate size

from the activity of reading text. Gaze can also move slowly in a 'smooth pursuit' manner in which the eyes 'lock on' to a moving image and track it.

When our gaze is directed to a particular location within a display the greatest detail is registered on the fovea, which is a small area in the centre of the retina. It is here, within a subtended angle of about two degrees, that vision is sharpest; at the periphery of vision we can discern only large objects. Another aspect of human visual performance is pupil diameter. It is widely believed that this measure is strongly influenced by various aspects of cognitive and perceptual behaviour.

Investigations of eye-gaze performance during the search for a target image among a collection of images is made possible by gaze detectors such as the one shown in Figure 4.47. An infra-red laser beam is (safely) directed towards each eye and the resulting reflections from retina and cornea are detected by a video camera and processed to identify the location of eye-gaze.

GURE 4.47 piece of uipment for e recording of e-gaze. An fra-red laser am is aimed the user's e and flections from e retina and rnea are tected by a levision mera. It also cords pupil ameter

4.2.3 Presentation modes

Cooper *et al.* (2006) investigated the relative advantages of the three static modes and three moving modes shown in Figure 4.45. Subjects were asked to say whether a target image[1] was present or not in a given presentation of 48 images and also to identify their most liked and least liked modes. The results were revealing. Figure 4.48 shows that all static modes were better than all moving modes with respect to the accuracy with which the presence or absence

[1] Three types of target image description were employed. For a given presentation of a collection of images, subjects (1) were shown the target image for a few seconds, or (2) had the target image described in specific terms (e.g. 'a cat'), or (3) had the target described in general terms (e.g. 'an animal'). Three total presentation times were also employed to see how this influenced the outcome of the experiment. The usual precautions were also taken to avoid well-known effects such as attentional blink (Raymond *et al.*, 1992). For more detail see Cooper *et al.* (2006).

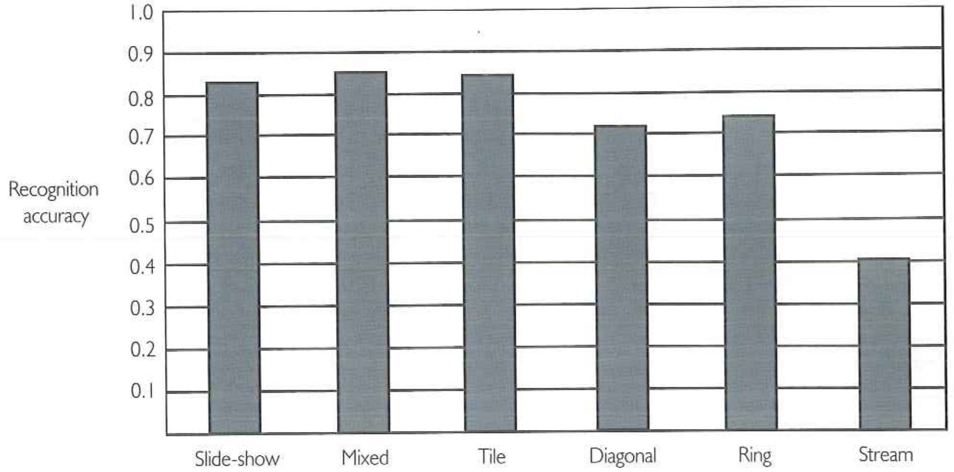

FIGURE 4.48 The accuracy with which the presence or absence of a target image was reported for the six presentation modes of Figure 4.45, averaged over all tasks and presentation times

of target images was identified, the stream mode being noticeably worst. Figure 4.49 also shows that the static modes (slide-show, mixed and tile) accounted for 75 per cent of preferred modes and only 25 per cent of subjects preferred a moving mode (though none preferred the stream mode). When asked to identify their least preferred mode (Figure 4.50), most identified a moving mode, and of those the stream mode accounted for many of the least preferred moving modes.

Eye-gaze data recorded during this and an earlier (Fawcett *et al.*, 2004) experiment is particularly revealing. Figure 4.51 shows some representative trajectories (saccades are green and fixations are denoted by 'F'). In a slide-show presentation (a) it is not surprising to find that gaze appears to be concentrated around the centre of the image. However, while eye-gaze behaviour for the tile mode (b) is again as expected due to the need to search for a target image, relatively little movement is recorded for the mixed mode (c): in some way the user appears to be able to assess all four images without undue movement, a fact that may account for that mode being highly preferred and characterized by high

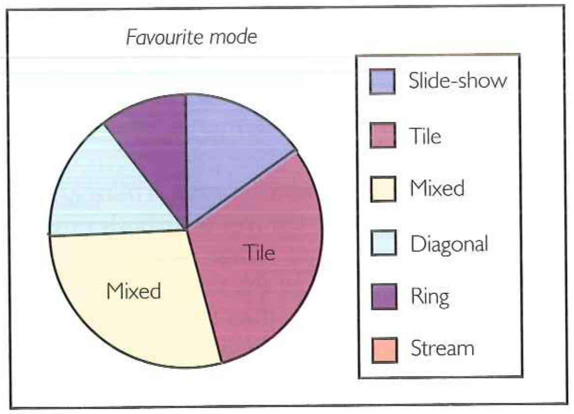

FIGURE 4.49 The (static) slide-show mixed and tile image presentation modes account for three-quarters of the preferred modes

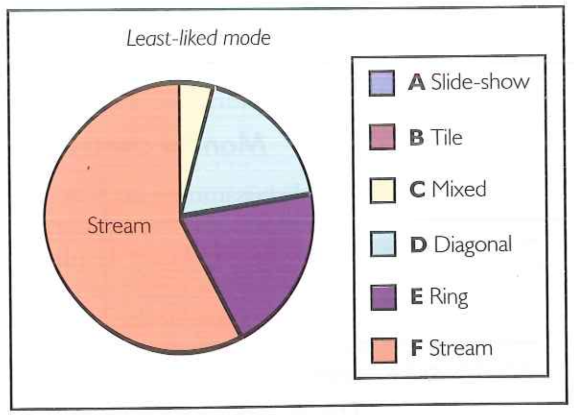

FIGURE 4.50 Almost all the least preferred image presentation modes were moving modes and the stream mode accounted for over half

accuracy of target identification. We turn now to the moving modes. Figure 4.51(d) shows the eye-gaze trace for a subject who disliked the diagonal (moving) mode, whereas (e) shows the trace for a subject who preferred it. The difference is striking and suggests that eye-gaze strategies adopted by subjects may underlie not only preferences but also target identification performance. Notable is the concentration of gaze at the 'capture' location (bottom right) in Figure 4.51(e) where, for a short while, an image is static. The subject is essentially viewing the image in slide-show (static) mode. It will not be surprising to learn that the trajectory shown in Figure 4.51(f) is that of a subject who disliked the stream mode.

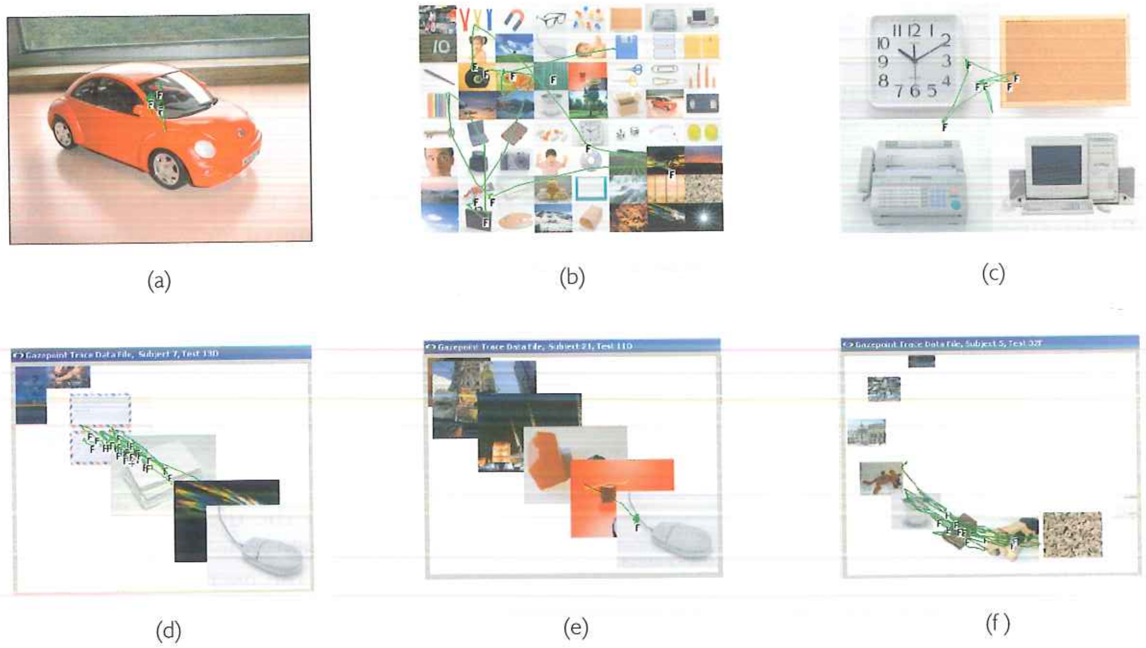

FIGURE 4.51 Representative eye-gaze trajectories recorded for selected image-collection presentation modes: (a) slide-show, (b) tile,(c) mixed, (d) diagonal, for a subject who disliked that mode, (e) diagonal, for a subject who liked that mode, (f) stream, for a subject who disliked that mode

4.2.4 *Manual control*

The discussion so far has assumed a fixed image presentation rate. There are circumstances, however, when a user might wish to exert control over both rate and direction, the latter to allow, for example, retreat to an image thought likely to have been the target or one that could now be defined as the target (e.g. a box of chocolates for Mother's Day). Not surprisingly many schemes for manual control have been proposed. We have already encountered manual control within 'floating RSVP', illustrated in Figure 4.40, and within 'collage RSVP', shown in Figure 4.41. The Nokia mobile photo library illustrated in Figure 4.16 can be controlled manually, as can the conventional 'scrolled tile' mode typified by iPhoto.

Another manually controlled presentation mode can be based on expanding images, a feature incorporated in the MacOSX tool bar and illustrated in Figure 4.52. In its dormant state the collection of images can be quite compact, with

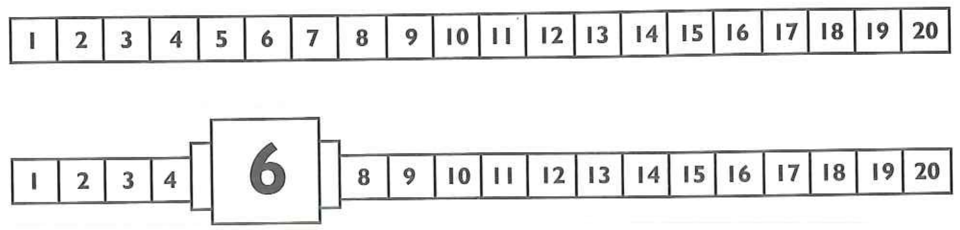

FIGURE 4.52 The acquisition of an expanding target. (a) The dormant appearance of the image collection, and (b) its appearance when the cursor rests over image 6

each image displayed in miniature so that the collection occupies little display area. Compaction may also be achieved by overlap. However, when a cursor is moved along the collection, individual images are considerably magnified, allowing a user to decide whether that image is a target. This 'expanding target' presentation mode has the potential advantage that little display area is needed to accommodate either the dormant or sampled image collection. In the extreme, this presentation technique closely resembles a manually controlled slide-show presentation. Indeed, recalling the experimental results reported for static and moving modes, one wonders if a 'snap-to-maximum-size' action, rather than a continuous expansion of image size, might lead to improved image recognition and enhanced preference. The acquisition of expanding targets has been investigated in depth by McGuffin and Balakrishnan (2002).

It is certainly the case that the design of any scheme allowing either the manually controlled or automated presentation of a set of images is a complex undertaking. The results concerning eye-gaze discussed above may be useful in both these circumstances. The presentation of image collections to support a variety of tasks continues to be the subject of intensive research (see, for example, Cockburn *et al.*, 2006).

4.2.5 *Models of human visual performance*

Though much experience has been accumulated regarding image presentation modes, our common understanding of human visual performance – which has a much wider relevance to information visualization as a whole – often lags behind. Nevertheless, it is pertinent to add some brief remarks here. For example, the experiment (Potter and Levy, 1969) illustrated in Figure 4.43 might suggest that, without seeing an initial target image, a user might still remember the general nature of ten or more rapidly presented images sufficiently to be

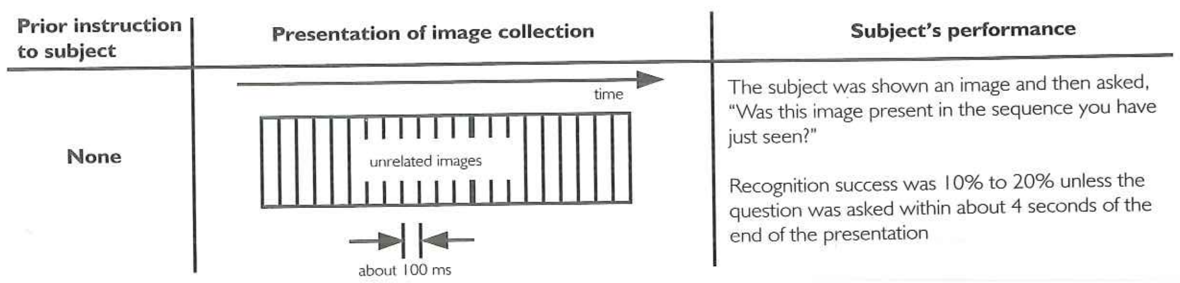

FIGURE 4.53 An experiment in which a subject first views the rapid (e.g. 10 per second) presentation of a collection of images, is then shown a single image and then asked to say whether that image was part of the collection. Identification success is highly dependent upon the time elapsing between the end of the presentation and the questioning of the subject