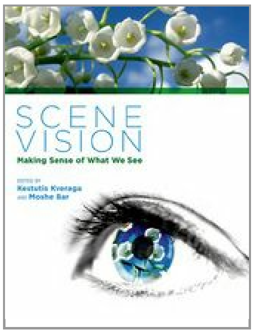

Scene Vision: Making Sense of What We See

Kestutis Kveraga and Moshe Bar

Print publication date: 2014
Print ISBN-13: 9780262027854
Published to MIT Press Scholarship Online: May 2015
DOI: 10.7551/mitpress/9780262027854.001.0001

Detecting and Remembering Briefly Presented Pictures

Mary C. Potter

DOI:10.7551/mitpress/9780262027854.003.0009

[–] Abstract and Keywords

This chapter proposes that conceptual information is extracted early in the visual process and forms what people remember later. By examining continual shifts of fixation by using rapid series visual sequences of unrelated pictures, the discussion shows that initial memory for a briefly presented picture is relatively accurate but declines slowly as time goes by. When unrelated pictures are presented in a continuous sequence at rates in the range of eye fixations, the memory for pictures is poor, suggesting that one glimpse is not sufficient for later memory. In contrast, pictures viewed for 1 to 10 seconds are more easily remembered. Accurate visual information may be important for maintaining scene representations from one thing to another but conceptual memory appears to be the basis for long-term knowledge.

Keywords: conceptual information, initial memory, conceptual memory, long-term knowledge, eye fixation

During our waking hours we take a new mental snapshot—a fixation—about three times a second. What do we pick up from each glimpse, and for how long do we remember what we saw? What is the form of our memory representation—visual, conceptual, or both—and does it change over time? One method for addressing these questions in the laboratory is to simulate continual shifts of fixation by using rapid serial visual presentation (RSVP) of sequences of unrelated pictures. When viewers are given a target name such as *picnic* or *smiling couple*, they are able to detect a picture in a stream presented for 100 ms per picture, and they do better than chance even at 13 ms/picture. Remarkably, detection is possible even when the name is given only after the sequence has been viewed. These results indicate that understanding may be based initially on feedforward processing, without feedback and without requiring advance information about the target. In contrast to our very rapid comprehension of pictures, we have poor memory for pictures presented for the duration of an average fixation (250 ms). We need 500 ms to view or think about a scene in order to remember it later. Yet long-term memory for pictures viewed for 1 second or more is excellent. The evidence suggests that conceptual information is extracted early and shapes what we remember later.

The paradox of vision is that we make three or four eye fixations each second, all day long, but each glimpse of 250 or 300 ms is too brief to remember later. We need some form of visual short-term memory that spans several fixations to integrate information about the immediate environment, but what we carry over from the preceding fixation lacks detail (e.g., Henderson & Hollingworth, 1999; Irwin, 1992; Irwin & Andrews, 1996). Moreover, studies of change blindness and boundary extension show that we overlook major changes in a scene if the scene is interrupted for as little as 80 ms (e.g., Intraub & Richardson, 1989; Rensink, O'Regan, & Clark, 1997, 2000), suggesting that our immediate memory is incomplete. We do notice changes that affect gist or changes to objects that we are attending or are about to fixate. Thus, the information that we carry over from the previous fixation seems to be meaningful rather than purely visual. However, when unrelated pictures are presented in a continuous **(p.178)** sequence at rates in the range of eye fixations (Potter & Levy, 1969), our memory for pictures is poor, implying that one glimpse is not sufficient for later memory.

In contrast, we have good long-term memory for pictures viewed for 1-10 seconds (Nickerson, 1965; Potter & Levy, 1969; Shepard, 1967; Standing, 1973). Pictures viewed for 3 seconds are remembered in detail, whether they represent single objects (Brady, Konkle, Alvarez, & Oliva, 2008) or complex scenes (Konkle, Brady, Alvarez, & Oliva, 2010).

Just How Quickly Do We Understand a Pictured Object or Scene?

Reaction Time

One answer to the question of how quickly a picture is understood is the reaction time (RT) to make a recognition response to a picture. Naming the picture is one such response, but that includes time to retrieve the name after one has already recognized what the object is, and even well-known names take time to retrieve: average RT for naming a familiar object is over 900 ms. A measure of understanding that does not require name retrieval is the time to decide whether the scene or object is a member of a category such as *animal*, This yes-no category detection task turns out to be considerably faster (a mean of about 600 ms) than the time to name a picture (Potter & Faulconer, 1975). These RT measures include the time for the information to pass from the retina to the visual cortex as well as decision and response processes that occur after identification (e.g., Potter, 1983). Research using a go/no-go response gives shorter responses in such category-detection tasks (see the review by Fabre-Thorpe, 2011). Of particular interest is the minimum RT at which performance is above chance, which has been shown to be as short as 150 ms. A still faster response is the initiation of an eye movement to a specified target (e.g., *animal* or *face*) when two pictures are presented simultaneously (e.g., Crouzet, Kirchner, & Thorpe, 2010; Kirchner & Thorpe, 2006): the shortest RT at which performance is above chance can be as little as 100 ms for faces, with a mean time of 140 ms. Another approach is to use measures of brain responses such as event-related potentials (ERPs) that occur before any overt response. In an early go/no-go study in which observers detected animals, the relevant ERP signal was significantly above chance beginning about 150 ms after picture onset (e.g., Thorpe, Fize, & Marlot, 1996).

Masked Stimuli

A different approach to measuring the time to understand a picture is to control the time available for processing the stimulus, measuring the minimum presentation time required to identify it.[1] However, the duration of the physical stimulus is not the same as the effective duration of the stimulus because of visual persistence: a picture presented for only 20 ms followed by a blank screen will persist for 80 ms or more. **(p.179)** A common method to solve that problem is to use a backward pattern mask at a variable delay after the picture (the stimulus onset asynchrony, SOA). Such a mask interrupts processing of the picture, allowing one to determine the minimal viewing time required for identification. For example, in one study (Potter, 1976; see figure 9.1, discussed below), 16 single pictures were each followed by a visual mask with an SOA varying from 50 to 120 ms. In a subsequent yes-no test of recognition memory about half the pictures were remembered at an SOA of 50 ms, rising to 80% at 120 ms.

Questions about Masking

A continuing problem with the logic of the masking procedure, however, is that the neural basis for the effect is not well understood: does the masked stimulus continue to be processed, perhaps unconsciously, after the mask appears, or does processing instantly stop? Macknik and Martinez-Conde (2007) have argued that the mask has an immediate feedforward effect that interrupts processing. But because the extent of masking depends not only on the SOA but also on the stimulus termination asynchrony and the perceptual relation of the mask to the stimulus of interest, the minimal SOA required for identification may not directly measure the time to understand a picture. Moreover, the effect of a following mask also depends on its semantic (conceptual) relation to the target picture. With very short SOAs the visual relation may be the major determinant of the mask's effectiveness, but as the SOA increases, the conceptual relation may be more important, as discussed below.

Context Effects: Perception of Objects and Settings

The role of visual context in perception of objects has long been a topic of interest. A similar question is whether our experience of co-occurrences between objects and settings influences the initial perception of a scene or whether (as suggested by Hollingworth & Henderson, 1998, 1999) objects and settings in a given picture are first understood independently. In one set of studies by Davenport and Potter (2004) pictured objects such as a football player or a priest were superimposed, either congruently or incongruently, on background settings such as a football field or the interior of a cathedral. The pictures were presented for 80 ms with a backward noise mask of the whole picture; the participant was instructed to report the foreground object, the background setting, or both. In each case performance was better in the congruent than the incongruent condition, suggesting that objects and background are processed interactively. When there were two objects in a scene, the likelihood that the two objects would be found together also influenced the report of the objects, an effect that was additive with the effect of congruency with the background (Davenport, 2007). Joubert and colleagues carried out similar studies, finding that objects in congruent contexts were responded to faster than those in incongruous contexts (Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007; Joubert, Fize, Rousselet, & **(p.180)** Fabre-Thorpe, 2008; see also Munneke, Brentari, & Peelen, 2013). These results indicate that objects and settings are processed together.

Rapid Serial Visual Presentation

In studies that use backward masking to limit processing time, each trial typically consists of a single stimulus, such as a picture, followed by a mask. Although the glimpse of the stimulus may be of the same duration as a fixation, in normal vision the eyes make a continuous sequence of fixations, with each fixation presumably masking the previous one. To mimic this effect Potter and Levy (1969) used a method called rapid serial visual presentation (RSVP) (Forster, 1970) to present pictures in a continuous stream at durations in the range of eye fixations, 125-2000 ms/picture. Participants were instructed to attend to and remember all the 16 pictures in a sequence. The pictures were unrelated to each other to enable us to measure memory for information equivalent to that in a single fixation. To test recognition memory following the presentation, the pictures were shown one at a time intermixed with 16 new pictures (distractors). Participants responded yes, maybe, or no. Figure 9.1 shows the proportion of yes responses, corrected for guessing.[2] When the pictures had been shown for the duration of an average fixation, 250 ms, fewer than half the pictures were correctly recognized a minute or two later. With a presentation of 2 seconds,

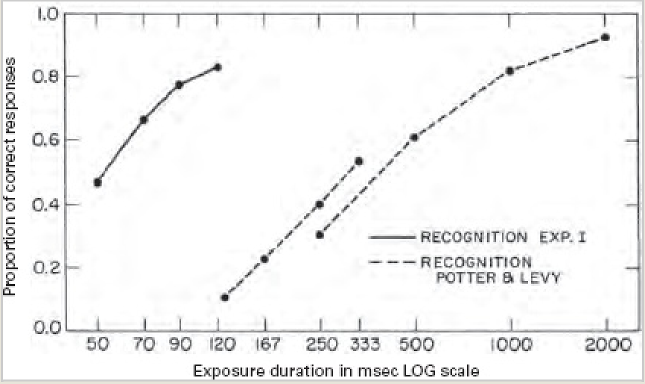

Figure 9.1 Proportion of pictures recognized following single masked presentations (solid curve, Potter, 1976, experiment 3) and proportion recognized after RSVP (dashed curves, two groups with different ranges of presentation durations; Potter & Levy, 1969). Data are corrected for guessing (see note 2). From Potter (1976).

(p.181) more than 90% of the pictures were remembered, consistent with studies showing that long-term memory for pictures viewed for a few seconds is excellent (e.g., Konkle et al., 2010; Nickerson, 1965; Shepard, 1967; Standing, 1973).

Visual versus Conceptual Masking

Strikingly, however, as shown in the left-hand function in figure 9.1 (Potter, 1976), a single masked picture may be remembered after it is viewed for as little as 50 ms (about 50% were remembered, rising to 80% at 120 ms). It takes four or five times as long, per picture, to process pictures to the same level of accuracy when they are presented in a continuous stream in which all the pictures are to be attended. Pictures in an RSVP sequence are hard to remember not only because of their briefness but also because each picture is immediately followed by another. With a single masked picture, viewers can continue to think about what they saw after the mask appears; that is not possible with a continuous sequence in which all the pictures are potentially relevant. In a study by Intraub (1980) pictures were presented for 110 ms in an RSVP sequence, and only 20% were remembered later, whereas when a blank interstimulus interval (ISI) was added after each picture, the percentage remembered increased steadily as the ISI increased, to 84% with an ISI of 1390 ms. Thus, a viewer can voluntarily continue to process and code into memory a brief picture after it is no longer in view, just as one can continue to think about what one just saw in a brief glimpse. Similarly, another study showed that pictures presented for 173 ms in an RSVP sequence were poorly remembered, but if a blank of 827 ms was added after each picture, memory was almost as good as if the pictures remained in view for the full 1000 ms (Potter, Staub, & O'Connor, 2004).

Voluntary Attention

In a study of the effect of voluntary attention on picture encoding, Intraub (1984) showed a sequence of pictures that alternated between a short duration of 112 ms and a long duration of 1500 ms. When viewers were instructed to attend to all pictures, they remembered about 54% of the short pictures and 73% of the long ones, whereas when instructed to attend only to the brief pictures, they remembered about 63% of the brief pictures and only 54% of the long ones. Altogether, these studies suggest that once the SOA between the picture and the following visual mask is 100 ms or more, memory depends little on the actual duration of presentation but instead on the total time the viewer continues to think about the picture. These results reinforce the distinction between visual and conceptual masking. Visual masking occurs primarily with short SOAs (under 100 ms), whereas conceptual masking (due to attention to a following stimulus) occurs with SOAs up to 500 ms or more (Potter, 1976' see also Intraub, 1980, 1981; Loftus & Ginn, 1984; Loftus, Hanna, & Lester, 1988; Loschky, Hansen, Sethi, & Pydimarri, 2010).

(p.182) Rapid Memory Loss for Pictures Seen Briefly in RSVP: Serial Position Effects in Memory Testing

People can understand pictures presented briefly but forget most of them a few minutes later. When the recognition test begins immediately, the first one or two pictures tested are likely to be remembered well, but there is rapid loss over the next several seconds of testing (Endress & Potter, 2012; Potter et al., 2004; Potter, Staub, Rado, & O'Connor, 2002); that is, there is a strong serial position effect in the memory test. There is also some loss if there is an unfilled delay of 5 seconds in the start of the memory test, showing that the loss is partly due to the passage of time and partly to interference from testing. Surprisingly, there is no serial position effect in presentation, apart from the known benefit to the final picture, which is unmasked and is not tested. Even with sequences as long as 20 items, there were no primacy or recency effects (Potter et al., 2002). Increasing the memory set size did decrease the extra benefit of early testing somewhat, but not by causing selective forgetting of pictures early in the list.

What Is the Nature of This Short-Lasting Memory for Pictures?

Change Blindness

The time course of forgetting after viewing an RSVP sequence of pictures contrasts with that of *change blindness*, the apparently immediate loss of detailed information about a single picture once it is no longer in view. Change blindness is the inability of viewers to detect a change in one feature of a picture, and the effect has been observed when a blank interval as short as 80 ms intervenes between the initial and changed versions; at longer intervals, the problem is even more acute (see Rensink et al., 1997; 2000; Simons & Levin, 1997). (Imposing a short blank between views is necessary to obscure the transient that would mark the location of the change if there were no interval.) Change blindness can be explained in several ways: the changed details were not perceived in the first place; many specifics of a picture are forgotten immediately; or the next picture updates the similar preceding picture without leaving a record of the changed details. Change blindness is, however, a very different phenomenon than the forgetting observed after an RSVP sequence. Whereas on a change blindness trial there is no question that the picture remains the same in most respects and is thus seen as the same picture, in the RSVP experiments considered here the question is whether a given test picture is one you have ever seen before. Thus, change blindness studies assess the level of detail in immediate memory for a picture, whereas here we are interested in the persistence of a representation sufficient to make the picture as a whole seem familiar.

(p.183) Other Forms of Brief Visual Memory

Could the short-lasting memory for pictures be *iconic memory* (e.g., Sperling, 1960) or *visual short-term memory* (VSTM) as described by Phillips and his colleagues (Phillips, 1983; Potter & Jiang, 2009)? The answer is, no. Iconic memory is a very brief form of relatively literal perceptual memory (although see Coltheart, 1983, for a somewhat different characterization), but it cannot account for the fleeting picture memory found with an immediate recognition test after an RSVP sequence because iconic memory is eliminated by noise masking, and under photopic conditions it lasts no longer than about 300 ms. VSTM is a form of short-lasting visual memory observed in experiments such as those of Phillips and Christie (1977), who presented viewers briefly with a 4 × 4 matrix in which an average of eight random squares were white and then tested memory by presenting a second matrix that was either identical to the preceding one or had one white cell added or deleted. VSTM, unlike iconic memory, is capacity limited, with an estimated capacity of three or four items. In Phillips and Christie's study the most recent matrix could be maintained for several seconds in VSTM provided that no other such matrices were presented in the interval and the participant continued to attend to the remembered matrix. In contrast, in RSVP studies multiple pictures are presented, and one or more to-be-attended pictures intervene between presentation and testing.

A likely contributor to short-term memory for pictures is *conceptual short-term memory* (CSTM), a short-lasting memory component proposed by Potter (1993, 1999, 2010) that represents conceptual information about current stimuli, such as the meaning of a picture or meanings of words and sentences computed as one reads or listens. The reasons for regarding this brief memory representation as conceptual rather than (say) perceptual include its apparent role in rapid selection between two words on the basis of meaning in relation to context (Potter, Moryadas, Abrams, & Noel, 1993; Potter, Stiefbold, & Moryadas, 1998) and its putative role in sequential visual search tasks like those considered here in which the targets are defined by meaning or category rather than by physical form. During the brief time that information about stimuli is in CSTM, associative links enable extraction of whatever structure is present (such as sentence structure or the gist of a picture) or allow the stimulus to be compared to a target specification in a search task. Any momentarily active information that does not become incorporated into such a structure (such as the irrelevant meaning of an ambiguous word or a nontarget picture) will be quickly forgotten.

Conceptual versus Visual-Perceptual Memory

A critical question is whether the picture representation that persists for several seconds in the studies we have reviewed here is sufficiently abstract to be considered conceptual rather than wholly or partly perceptual. Do viewers remember only the **(p.184)** picture's conceptual content or gist, or do they also remember visual features such as color, shape, and layout? Work of Irwin and Andrews (1996), Gordon and Irwin (2000), and Henderson (1997) suggests that the representation of the previous fixation may be at least partially conceptual rather than literal inasmuch as viewers may not notice literal changes that are conceptually consistent with the earlier fixation. Studies of detection to be reviewed below show that the gist of a scene is understood quickly even though the scene may then be forgotten (fairly) rapidly (e.g., Intraub, 1980, 1981; Potter, 1976), which is consistent with the assumption that conceptual information is abstracted rapidly. Intraub (1981) showed, however, that viewers can remember some specific pictorial information, such as the colors and layout, along with the gist.

The relative roles of such specific pictorial information and more abstract conceptual information were explored by Potter et al. (2004). They contrasted a conceptual and a pictorial recognition test of picture memory. In the pictorial test participants made yes-no decisions to five pictures they had just seen (excluding a sixth final picture that was not masked), mixed with five new pictures. In the conceptual test they made yes-no decisions to descriptive verbal titles of the pictures, mixed with titles of unseen pictures. The presentation duration was 173 ms/picture; the 10-item recognition test after each trial took about 8 seconds. The assumption was that test pictures provide both visual and conceptual information, whereas titles provide only conceptual information. If the benefit of immediate testing is that viewers only briefly preserve purely pictorial information, then the title test should reduce the benefit of early testing but should be fairly equivalent to the picture test later in testing. That was just what they found. In a more recent study (Endress & Potter, 2012) the advantage of testing recognition with pictures rather than titles was maintained throughout the test, suggesting that some more detailed information (perceptual or conceptual) beyond that captured by a title does persist over the 8-second test even though memory for both forms of information continues to decline.

In a further test of the conceptual basis of memory, Potter et al. (2004) included in the recognition test occasional *decoy* pictures that matched the title—the gist—of one of the old pictures, replacing that picture in the test. The decoy looked visually different from the old picture it replaced. If viewers rely on a conceptual or gist representation of the presented pictures, they should make more false yeses to decoys than to unrelated new pictures (distractors). Overall, participants recognized 52% of the old pictures, falsely recognized 30% of the decoys, and falsely recognized 15% of the other distractors, showing some susceptibility to conceptual decoys.

Short-Lasting Memory: Summary

Initial memory for a glimpsed picture (seen for the equivalent of a single fixation) is fairly accurate but declines markedly over the first few recognition tests (or across an **(p.185)** unfilled delay of 5 seconds). The initial stronger memory may include specifically visual information, whereas after a delay the memory is primarily conceptual. Accurate visual information may be important for maintaining and updating scene representations from one fixation to the next, but conceptual memory seems to be the basis for longer-term organized knowledge. Unlike the rapid forgetting of briefly glimpsed pictures, memory for pictures viewed for a second or more can be highly accurate, at least when viewers are paying attention.

Detecting Pictures to Test Comprehension

Are RSVP Pictures Understood?

The studies of picture memory that I have just reviewed show that pictures presented for durations in the range of typical eye fixations are not well remembered. How do we know whether the forgotten pictures were even understood momentarily? Subjectively, one has the impression that one understands all the pictures when presented up to 10/second, but perhaps that is an illusion. Does it take longer than a single fixation to understand a novel scene? Perhaps viewers fail to remember briefly presented pictures because they did not comprehend them. To discover whether brief pictures are identified but then forgotten, we asked participants to detect target pictures that were shown to them (or named) before the sequence (Potter, 1975, 1976). We used names that captured the conceptual gist of the picture in one to five words but did not give explicit visual information about the picture. Detection was surprisingly good with either kind of cue, even at durations as short as 113 ms/picture (figure 9.2). The results can be compared with the recognition memory results from another group who viewed the same sequences without looking for a target and whose members were tested after each sequence for their recognition memory. That group, also shown in figure 9.2, remembered far fewer pictures than the first group had detected, suggesting that viewers can momentarily understand most of these brief pictures but will then forget many of them as testing begins. A question arises, however, about whether the difference between the two groups in figure 9.2 simply reflects attentional set: having a name presets the visual system to process the scene, which would not be understood otherwise. Intraub (1981) addressed that question by showing that viewers could detect a picture described by a negative category such as “not an animal,” although performance was not as good as when given the name. The role of attentional set is considered again below, when we consider the difference between naming a picture before versus immediately after the sequence.

Detecting Two Targets

In another detection study (Potter, Wyble, Pandav, & Olejarczyk, 2010), participants looked for two targets in a category such as “bird” and reported the specific identity **(p.186)**

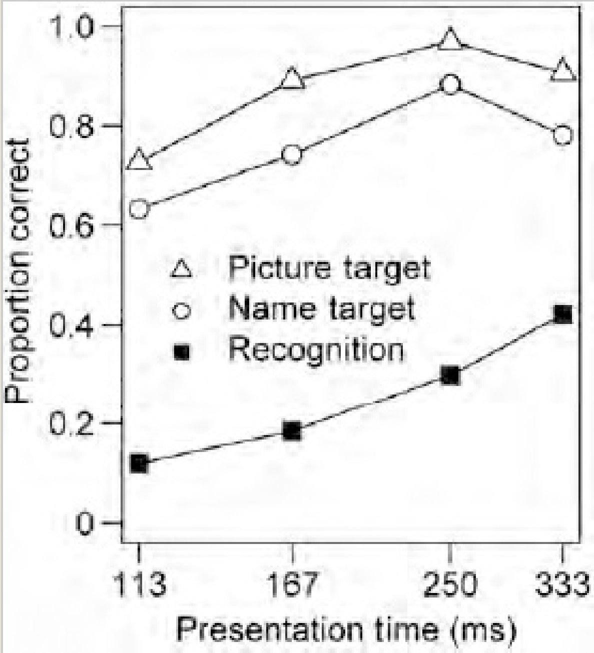

Figure 9.2 Detection of a target picture in an RSVP sequence of 16 pictures, given a picture of the target or a name for the target, as a function of the presentation time per picture. Also shown is later recognition performance in a group that simply viewed the sequence and then was tested for recognition. Results are corrected for guessing (see note 2). From Potter (1976).

of each instance (e.g., *swan* and *eagle*). The RSVP sequence was shown at 107 ms/picture. Figure 9.3 illustrates a trial in which the category was "dinner food." Report of the specific names of both targets (e.g., hamburger, spaghetti) was often successful even when the two targets were presented in immediate succession, although there was an attentional blink (reduced performance) for the second target when the SOA between targets was 213 ms, an effect typically observed in search tasks. Thus, even when given a general name for the target, viewers could detect and retain the specific identities of two targets presented briefly in a sequence.

Detection and Memory when Multiple Pictures Are Presented Simultaneously

Potter and Fox (2009) presented eight successive four-item arrays (figure 9.4) in which each array included none to four pictures, with meaningless texture masks filling the nonpicture locations. The RSVP sequence was presented at 240, 400, or 720 ms per array. When the task was to detect a named target (e.g., *balloons*), detection was relatively successful with up to four simultaneous pictures. Even at 240 ms per array with four simultaneous pictures, 59% of the targets were detected, with 9% false yeses (cf. Rousselet, Fabre-Thorpe, & Thorpe, 2002; Rousselet, Thorpe, & Fabre-Thorpe, 2004a, 2004b). This suggests that detection occurs in parallel with up to four pictures, or detection is extremely fast, or both. When viewers simply tried to remember the **(p.187)**

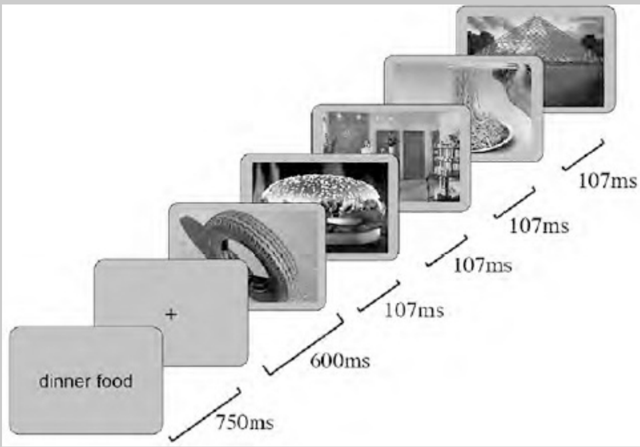

Figure 9.3 An example of an RSVP sequence in a search experiment in which participants reported the specific names of two exemplars of the search category. Here the exemplars are *hamburger* and *spaghetti*. From Potter, Wyble, Pandav, and Olejarczyk (2010).

pictures, later recognition accuracy was much lower overall, particularly when there was more than one picture in the array. We speculate that detection may occur in parallel over the whole array, whether it consists of one picture or up to four. In contrast, memory may require separate attention to each picture.

Detection and Memory with Occlusion, Inversion, and Grayscale Pictures

Meng and Potter (2008) used RSVP to present pictures with or without 30% of the surface randomly occluded by small disks and found that detection (given a name) was well above chance and minimally affected by the disks, even with a duration as brief as 53 ms. When the task was to recognize a picture shown after the sequence, performance was lower than with detection, and the disks significantly interfered. When the pictures were inverted, the disks interfered with detection as well as recognition. Showing the pictures in grayscale did not change performance in the detection condition, and again the occluding disks did not affect performance. When the number of disks was increased to cover 40% of the picture, however, detection did show interference. The results suggest that rapid retrieval of the gist of a picture is based on a global perception of the scene that is robust against local loss of information. **(p.188)**

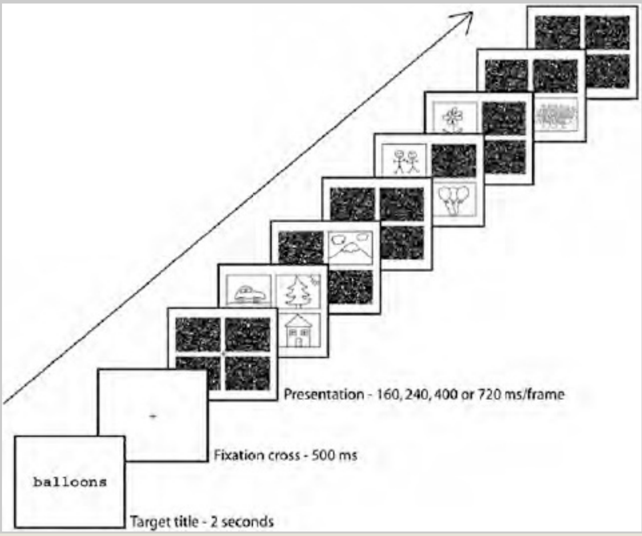

Figure 9.4 Schematic representation of an RSVP sequence in which up to four pictures were presented simultaneously for detection of one named target. From Potter and Fox (2009).

Detecting Pictures at Ultrahigh Rates: Evidence for Feedforward Processing?

Feedforward Processing

In feedforward models of the visual system (Serre, Kreiman, et al., 2007; Serre, Oliva, & Poggio, 2007), units that process the stimulus are hierarchically arranged. Units representing small regions of space (receptive fields) in the retina converge to represent larger and larger receptive fields and more abstract information along a series of pathways from V1 to inferotemporal cortex (IT) and further on to prefrontal cortex (PFC). Visual experience tunes this hierarchical structure, which acts as a filter that permits recognition of a huge range of objects and scenes in a single forward pass of processing. Yet, there is little direct evidence that the feedforward process is able to identify objects and scenes accurately, without feedback. Under normal viewing conditions perception is generally assumed to result from a combination of feedforward and feedback connections (DiLollo, Enns, & Rensink, 2000; Enns & Di Lollo, 2000; Hochstein & Ahissar, 2002; Lamme & Roelfsema, 2000). Feedback from higher to lower levels in the visual system takes time, however. At presentation durations of **(p.189)** about 50 ms or less with masking, some have proposed that there would not be time for feedback to arrive before the lower-level activity has been interrupted by the mask, so that perception, if any, would be restricted to the information in the forward pass of neural activity from the retina through the visual system (Hung, Kreiman, Poggio, & DiCarlo, 2005; Liu, Agam, Madsen, & Kreiman, 2009; Perrett, Hietanen, Oram, & Benson, 1992; Thorpe & Fabre-Thorpe, 2001).

Conscious Perception

The ability to identify or remember a stimulus is commonly taken to mean that the viewer was conscious of the stimulus, and here I make the assumption that consciousness is shown by the ability to report on the stimulus by responding to a target picture or by recognizing its title or the picture itself in a memory test. (See, however, evidence for unconscious effects, discussed below.) There is a debate about whether a single forward pass is sufficient for conscious perception. A reentrant process providing feedback may be necessary to achieve understanding and conscious awareness (Dehaene & Naccache, 2001; Hochstein & Ahissar, 2002; Lamme & Roelfsema, 2000). As mentioned above it has been suggested that a threshold duration of about 50 ms must be exceeded if there is a backward mask, or the stimulus will not be consciously perceived. Consciousness of a stimulus may require sufficient time “to establish sustained activity in recurrent cortical loops” (Del Cul, Baillet, & Dehaene, 2007) or to ignite a network required for conscious perception (Deheane, Kergsberg, & Changeux, 1998). These authors thus hypothesize that viewers cannot become conscious of a stimulus on the basis of a single feedforward sweep, without time for any feedback. Detection in RSVP at durations of 50 ms per picture or less should be impossible if there is such a threshold because there is too little time to establish a long-range cortical loop before a picture has been overwritten by subsequent pictures. As reviewed in the next section, however, there is evidence that perception is sometimes possible with very brief masked stimuli, a result that suggests that feedforward processing may be sufficient for conscious perception under some conditions.

Evidence for Processing of Very Brief Stimuli: RSVP Responses by Monkey Neurons and Humans

Recordings of individual neurons in the cortex of the anterior superior temporal sulcus (STSa) of monkeys that viewed a set of pictures of monkey faces and other objects via RSVP at various rates up to 72 per second (14 ms) showed that neurons respond to a preferred picture above chance, even at 14 ms (Keysers, Xiao, Földiák, & Perrett, 2001, 2005). In a detection study with human observers using the same set of pictures but presenting them in seven-picture RSVP sequences, the participants were shown a target picture before each sequence. They detected the target above chance at 14 ms per picture, although detection improved as the duration per picture **(p.190)** was increased. In another condition in the same study, recognition of a target picture was tested immediately after the sequence, instead of being shown before the sequence. Participants were still above chance at 14 ms per picture, but performance was not as good as when they saw the target picture in advance. A possible problem with the human study is that the pictures were repeated across trials and hence became familiar, which might have allowed participants to focus on simple features in order to spot the target.

Detection and Immediate Memory for Conceptually Defined Targets

A study by Potter, Wyble, Hagmann, and McCourt (2014) replicated some of the behavioral conditions of Keysers et al. (2001), but crucially, instead of showing the picture target, they gave only a descriptive name for the target (e.g., *smiling couple*), before or immediately after an RSVP sequence of six pictured scenes (figure 9.5 shows the method). Moreover, each picture was presented only once, and none of the pictures was familiar to the participants. Thus, participants had only a conceptual representation of the target they were to detect or recollect. The RSVP sequence was presented at durations between 13 and 80 ms. Even at a presentation duration of 13

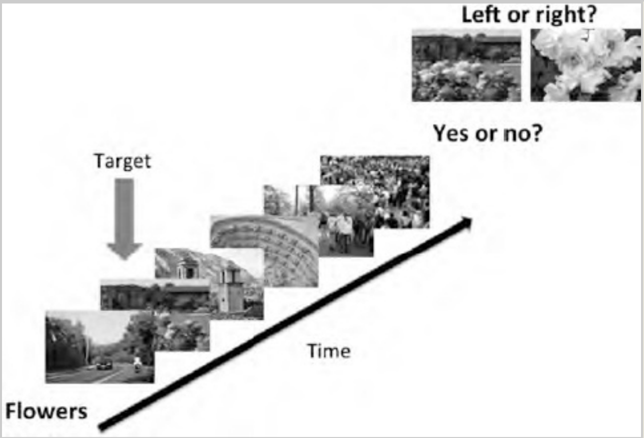

Figure 9.5 Illustration of a six-picture sequence in which the target is named in advance and a yes-no decision is followed by a forced choice between two pictures, both of which match the target name. See Potter, Wyble, Hagmann, and McCourt (2014).

(p.191) ms, the targets were detected or recognized above chance: that is, the probability of a correct detection on target-present trials was significantly higher than the probability of a false detection response on target-absent trials. In addition, at the end of each trial participants were shown two pictures, both matching the target name, and asked to indicate which one they had seen. They were above chance in selecting the right picture only if they had correctly detected the target; if they missed the target, their forced choice was at chance. Thus, viewers could detect and retain at least briefly information about named targets they had never seen before at an RSVP duration as short as 13 ms. A second experiment replicated those results with sequences of 12 rather than 6 pictures: again, detection and recognition were above chance at all durations, including 13 ms.

These results are consistent with the claim of the feedforward model that pictures can be understood in a single feedforward sweep even when attention has not been directed to a specific category in advance. In the name-after condition the participant had no knowledge of the target at the time he/she viewed the picture sequence, so the pictures had to be processed bottom-up and encoded. Only after the target was named could the participant search recent memory for the target—there was no top-down influence on perception, only on memory search. Performance was somewhat lower when the target name came after the sequence, rather than before, showing that advance information did make detection more likely.

Feedforward Processing and Masked Priming

In masked priming studies a brief presentation of a word becomes invisible when it is followed by a second unmasked word to which the participant must respond (Dehaene et al., 2001; Forster & Davis, 1984). If the prime word is related in some way to the following word, it may increase the accuracy or speed of response to the latter, showing that the prime must have been unconsciously identified. Given that the prime may have been presented for 50 ms or more in typical masked priming experiments (above the threshold for perception with a noise mask), why is the participant not conscious of the prime? In such studies the focus of attention is on the second stimulus, and its longer duration permits it to receive full, recurrent processing that may interfere with retention of the more vulnerable information from the prime that was extracted during the feedforward sweep. When, as in Potter et al. (2014) the masking stimulus has the same duration as the preceding target stimulus and is another picture that is to be attended, a duration of 13 ms is clearly sufficient, on a significant proportion of trials, to drive detection, identification, and (at least briefly) recognition memory for the pictures. It seems likely, however, that the reportable detection observed with RSVP tasks such as those described here has the same neural basis as masked priming.

(p.192) Discussion

Why are eye fixations so brief? It is clear from the research reviewed here that a typical single fixation of 250 ms is long enough to make it highly likely that the viewer will understand what he or she has looked at, at least momentarily. Yet, normal eye fixations are too brief to guarantee good memory. Why don't we fixate for longer? It appears that the rate at which we move our eyes is just slow enough to allow momentary understanding and to initiate appropriate action if needed (including taking a second look) but still fast enough to keep up with rapid changes in the scene around us, allowing us to dodge a bicycle or catch a ball. If something is important enough to need to be remembered, we can keep looking or keep thinking about it.

How long does it take to understand a pictured scene? To return to a question considered at the beginning of the chapter, what can be concluded about the time required to identify a scene? If the question is the minimum exposure duration (prior to a mask) that is required, 13 ms is sometimes enough when the mask is another scene. But if the question is the time from arrival at the retina to correct categorization, then the most reliable measures available at present are reaction time measures, the most sensitive of which is an eye movement to the appropriate target in a choice situation. For detection of a face (when a picture with a face is presented together with another picture), that time can be as short as 100 ms, with a mean time of 140 ms (Crouzet et al., 2010); detection of a vehicle takes somewhat longer. Momentary comprehension is no guarantee of subsequent memory, however, even seconds later. We comprehend rapidly, but then we forget selectively on the basis of what is relevant to our current goals and needs.

Ultrarapid Presentations and Feedforward Processing

Both the results of Keysers et al. (2001, 2005) with monkey neurons and with humans and those of Potter et al. (2014) with humans show that pictures can be detected and briefly remembered when presented in a short sequence at a rate as high as 75 pictures per second. Even when no target is specified in advance, a name presented immediately after the sequence can prompt memory for the corresponding picture. These results support a feedforward model that can extract a picture's conceptual meaning in a single forward sweep of information with an input as brief as 13 ms without requiring feedback loops from higher to lower levels and back and without requiring a selective attentional set. However, a longer feedforward viewing time of up to 80 ms may be required to grasp the gist of many scenes. When a scene is complex or its components are unfamiliar, we may need more than a single fixation to comprehend it.

(p.193) Although the results reviewed here indicate that feedforward processing is capable of activating the conceptual identity of a picture even when the picture is briefly presented and is then masked by immediately following pictures, they leave open the possibility that top-down or reentrant loops facilitate processing and may be essential to comprehend details. For example, there is evidence that a rapid but coarse first pass of low-frequency information may provide global category information that is subsequently refined by top-down processing (e.g., Bar et al., 2006). Other work has shown that monkey neurons that are selective for particular faces at a latency of about 90 ms give further information about facial features beginning about 50 ms later (Sugase, Yamane, Ueno, & Kawano, 1999), suggesting reentrant processing (DiLollo, 2012). In any case such reentrant theories rely on feedforward processing to generate tentative interpretations of a picture that are fed back and compared with the representations in earlier levels of processing, suggesting that feedforward processing initiates visual understanding.

But are there other explanations for successful detection when the presentation duration is brief and masked by successive pictures? One possibility is that subsequent pictures do not interrupt processing immediately. As mentioned earlier the neural basis for masking is not well understood. Studies of the monkey visual system using single-cell recordings show that multiple cortical neurons that are selective for different objects can be activated at the same time, suggesting that multiple objects may be "recognized" in parallel at levels as high as the inferior temporal cortex. Something similar in human perception might account for the ability to perceive rapidly presented pictures. In monkeys this initial parallel process is followed within 150 ms by competitive inhibition of all neurons other than those responding to the relevant object in a given receptive field, at least when there is a task that defines the relevant stimulus (e.g., Chelazzi, Duncan, Miller, & Desimone, 1998; see Rousselet et al., 2004a, for a review). The large and overlapping receptive fields found in the inferior temporal cortex may allow for temporary representation in parallel of several successive pictures presented at a high rate, followed by competitive suppression that favors the most salient picture. That could account for the ability on some trials to detect a target by name immediately after the presentation of 6 pictures (although with 12 pictures one would have expected a larger decrement than we observed). If high-level representations of several of the pictures in the sequence were activated, however, it is likely that mutual competition would soon decrease their activation well before the target name was presented. In further experiments with RSVP using very rapid sequences (Potter, Wyble, & Hagmann, unpublished data) a delay of 5 seconds in providing the target name after the sequence did decrease accuracy.

Thus, the feedforward hypothesis remains a strong contender as an explanation of picture identification with very brief presentation durations. In the absence of a specific model for how feedback might assist reportable detection of brief targets, the **(p.194)** feedforward hypothesis seems the most plausible account. A lifetime of experience of the world that is built into our visual system appears to allow immediate understanding of most scenes, based on the initial sweep of visual information when the scene is presented.

Acknowledgments

This work was supported by Grant MH47432 from the National Institute of Mental Health. I thank Carl Hagmann for comments and my many collaborators for their participation in the research reported here. Much of this work was also described in Potter (2012).

Note

References

Bibliography references:

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmidt, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America, 103*(2), 449—454.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America, 105*(38), 14325-14329.

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology, 80*(6), 2918-2940.

Coltheart, M. (1983). Iconic memory. *Philosophical Transactions of the Royal Society of London, 302*(1110), 283-294.

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades towards face: Face detection in just 100 ms. *Journal of Vision, 10*(4), 1-17.

Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory & Cognition, 35*, 393-401.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science, 15*(8), 559–564.

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition, 79*, 1-37.

Dehaene, S., Naccache, L., Cohen, L., LeBihan, D., Mangin, J. F., Poline, J.-B., et al. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience, 4*, 752-758.

Deheane, S., Kergsberg, M., & Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences of the United States of America, 95*, 14529-14534.

Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biology, 5*, 2408-2423.

DiLollo, V. (2012). The feature-binding problem is an ill-posed problem. *Trends in Cognitive Sciences. 16*(6), 317-321.

DiLollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual pathways. *Journal of Experimental Psychology. General, 129*.481-507.

Endress, A. D., & Potter, M. C. (2012). Early conceptual and linguistic processes operate in independent channels. *Psychological Science, 23*, 235-245.

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences, 4*, 345-352.

Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Psychology, 2*(243), 1-12.

Forster, K. I. (1970). Visual perception of rapidly presented word sequences of varying complexity. *Perception & Psychophysics, 8*(4), 215-221.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 10*, 680-698.

Gordon, R. D., & Irwin, D. E. (2000). The role of physical and conceptual properties in preserving object continuity. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 26*, 136-150.

Henderson, J. M. (1997). Transsaccadic memory and integration during real-world object perception. *Psychological Science, 8*(1), 51-55.

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

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36*(5), 791-804.

Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology. General, 127*(4), 398-415.

Hollingworth, A., & Henderson, J. M. (1999). Object identification is isolated from scene semantic constraint: Evidence from object type and token discrimination. *Acta*

Psychologica, 102(2-3), 319-343.

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science, 310*(5749), 863-866.

Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. *Journal of Experimental Psychology. Human Learning and Memory, 6*, 1-12.

Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology. Human Perception and Performance, 7*, 604-610.

Intraub, H. (1984). Conceptual masking: The effects of subsequent visual events on memory for pictures. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 10*(1), 115-125.

Intraub, H., & Richardson, M. (1989). Wide-angle memories of close-up scenes. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 15*(2), 179-187.

Irwin, D. E. (1992). Memory for position and identity across eye movements. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 18*, 307-317.

Irwin, D. E., & Andrews, R. V. (1996). Integration and accumulation of information across saccadic eye movements. In T. Inui & J. L. McClelland (Eds.), *Attention and Performance XVI: Information integration in perception and communication* (pp. 125-155). Cambridge, MA: MIT Press.

Joubert, O. R., Rousselet, G., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research, 47*, 3286-3297.

Joubert, O. R., Fize, D., Rousselet, G., & Fabre-Thorpe, M. (2008). Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *Journal of Vision, 13*, 11.

Keysers, C., Xiao, D. K., Földiák, P., & Perrett, D. I. (2001). The speed of sight. *Journal of Cognitive Neuroscience, 13*, 90-101.

Keysers, C., Xiao, D.-K., Földiák, P., & Perrett, D. I. (2005). Out of sight but not out of mind: The neurophysiology of iconic memory in the superior temporal sulcus. *Cognitive Neuropsychology, 22*, 316-332.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46*(11), 1762-1776.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science, 21*(11), 1551-1556.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23*, 571-579.

Liu, H., Agam, Y., Madsen, J. R., & Kreiman, G. (2009).Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron, 62*(2), 281-290.

Loftus, G. R., & Ginn, M. (1984). Perceptual and conceptual masking of pictures. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 10*, 435-441.

Loftus, G. R., Hanna, A. M., & Lester, L. (1988). Conceptual masking: How one picture captures attention from another picture. *Cognitive Psychology, 20*, 237-282.

Loschky, L. C., Hansen, B. C., Sethi, A., & Pydimarri, T. N. (2010). The role of higher order image statistics in masking scene gist recognition. *Attention, Perception & Psychophysics, 72*(2), 427-444.

Macknik, S. L., & Martinez-Conde, S. (2007). The role of feedback in visual masking and visual processing. *Advances in Cognitive Psychology, 3*, 125-152.

Meng, M., & Potter, M. C. (2008). Detecting and remembering pictures with and without visual noise. *Journal of Vision, 8*(9), 1-10.

Munneke, J., Brentari, V., & Peelen, M. (2013). The influence of scene context on object recognition is independent of attentional focus. *Frontiers in Psychology, 4*, 552.

Nickerson, R. S. (1965). Short-term memory for complex meaningful visual configurations: A demonstration of capacity. *Canadian Journal of Psychology, 19*, 155-160.

Perrett, D., Hietanen, J., Oram, M., & Benson, P. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 335*, 23-30.

Phillips, W. A. (1983). Short-term visual memory. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 302*, 295-309.

Phillips, W. A., & Christie, D. F. M. (1977). Components of visual memory. *Quarterly Journal of Experimental Psychology, 29*(1), 117-133.

Potter, M. C. (1975). Meaning in visual scenes. *Science, 187*, 965-966.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology. Human Learning and Memory, 2*(5), 509-522.

Potter, M. C. (1983). Representational buffers: The eye-mind hypothesis in picture perception, reading, and visual search. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 423-437). New York: Academic Press.

Potter, M. C. (1984). Rapid serial visual presentation (RSVP): A method for studying language processing. In D. Kieras & M. Just (Eds.), *New methods in reading comprehension research* (pp. 91-118). Hillsdale, NJ: Erlbaum.

Potter, M. C. (1993). Very short-term conceptual memory. *Memory & Cognition, 21*, 156-161.

Potter, M. C. (1999). Understanding sentences and scenes: The role of conceptual short term memory. In V. Coltheart (Ed.), *Fleeting memories: Cognition of brief visual stimuli* (pp. 13-46). Cambridge, MA: MIT Press.

Potter, M. C. (2010). Conceptual short term memory. *Scholarpedia, 5*(2), 3334.

Potter, M. C. (2012). Recognition and memory for briefly presented scenes. *Frontiers in Psychology, 3*, 32.

Potter, M. C., & Faulconer, B. A. (1975). Time to understand pictures and words. *Nature, 253*, 437-438.

Potter, M. C., & Fox, L. F. (2009). Detecting and remembering simultaneous pictures in a rapid serial visual presentation. *Journal of Experimental Psychology. Human Perception and Performance, 35*, 28-38.

Potter, M. C., & Jiang, Y. V. (2009). Visual short-term memory. In T. Bayne, A. Cleeremans, & P. Wilken (Eds.), *Oxford companion to consciousness* (pp. 436-438). Oxford: Oxford University Press.

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81*(1), 10-15.

Potter, M. C., Moryadas, A., Abrams, I., & Noel, A. (1993). Word perception and misperception in context. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 19*, 3-22.

Potter, M. C., Staub, A., & O'Connor, D. H. (2004). Pictorial and conceptual representation of glimpsed pictures. *Journal of Experimental Psychology. Human Perception and Performance, 30*, 478-489.

Potter, M. C., Staub, A., Rado, J., & O'Connor, D. H. (2002). Recognition memory for briefly-presented pictures: The time course of rapid forgetting. *Journal of Experimental Psychology. Human Perception and Performance, 28*(5), 1163-1175.

Potter, M. C., Stiefbold, D., & Moryadas, A. (1998). Word selection in reading sentences: Preceding versus following contexts. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 24*, 68-100.

Potter, M. C., Wyble, B., & Hagmann, C. E. (unpublished data). A delay of 5 seconds decreases *d'* in memory for rapid sequences.

Potter, M. C., Wyble, B., Hagmann, C. E., & McCourt, E. S. (2014). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, & Performance*.

Potter, M. C., Wyble, B., Pandav, R., & Olejarczyk, J. (2010). Picture detection in RSVP: Features or identity? *Journal of Experimental Psychology. Human Perception and Performance, 36*, 1486-1494.

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

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (2000). On the failure to detect changes in scenes across brief interruptions. *Visual Cognition, 7*, 127-145.

Rousselet, G., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high level categorization of natural images. *Nature Neuroscience, 5*, 629-630.

Rousselet, G., Thorpe, S. J., & Fabre-Thorpe, M. (2004a). How parallel is visual processing in the ventral pathway? *Trends in Cognitive Sciences, 8*(8), 363-370.

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004b). Processing of one, two or four natural scenes in humans: The limits of parallelism. *Vision Research, 44*(9), 877-894.

Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U., & Poggio, T. (2007). A quantitative theory of immediate visual recognition. *Progress in Brain Research, 165*, 33-56.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences of the United States of America, 104*(15), 6424-6429.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6*, 156-163.

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

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74*(498), 11.

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology, 25*, 207-222.

Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature, 400*, (869-873).

Thorpe, S. J., & Fabre-Thorpe, M. (2001). Seeking categories in the brain. *Science, 291*, 260-263.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520-522.

Notes:

(1.) A measure of the minimal time required for successful processing does not include the time for the retinal signal to arrive at the part of the brain doing the processing, which may take 60-80 ms, or the time to generate an overt response to the stimulus once it is understood (e.g., Potter, 1984).

(2.) A one-high-threshold formula was used to correct for guessing, Pcorr = [P(TY) - P(FY)]/[1 - P(FY)], where TY is a correct yes response and FY is a false yes response. This guessing correction is used in all data figures.