

FLEETING MEMORIES
Cognition of Brief Visual Stimuli

Edited by Veronika Coltheart

1999
The MIT Press
Cambridge, Massachusetts

Chapter 2

Understanding Sentences and Scenes: The Role of Conceptual Short-Term Memory

Mary C. Potter

In this chapter I summarize a thesis concerning a very brief form of memory that I have termed conceptual short-term memory (CSTM). I then consider the role of this form of working memory in attending selectively, understanding language, and recognizing and remembering scenes. Finally, I discuss some issues relevant to CSTM, including the question of whether CSTM is conscious.

When people view or listen to continuous sequences of scenes or words, as they do when they look around, read, listen, or watch TV, a series of conceptual representations is activated. These rapidly activated and equally rapidly forgotten representations are the raw material for identification and comprehension of words, pictures, and sequences such as a sentence, and indeed for intelligent thought more generally. The normal ease with which we understand what we read and see around us is based on selective processing that takes place much faster than has been supposed in many theories of working and short-term memory, leading to the CSTM hypothesis (Potter 1993).

The CSTM hypothesis proposes that when a stimulus is identified, its meaning is rapidly activated and maintained briefly in CSTM. CSTM is a processing and memory system different from early visual (iconic) memory, conventional short-term memory (STM), and longer-term memory (LTM) in three important respects: (1) the rapidity with which stimuli reach a postcategorical, meaningful level of representation, (2) the rapid structuring of these representations, and (3) the lack of awareness (or immediate forgetting) of information that is not structured or otherwise consolidated. Structure-building in CSTM ranges from spontaneous grouping of words in lists on the basis of meaning (one of the simplest forms of conceptual structuring) to linguistic parsing and semantic interpretation of sentences and more extended texts (examples of highly skilled structuring). Organization or structuring of new stimuli enhances memory for them.

The idea here is that most cognitive processing occurs on the fly, without review of material in standard short-term memory and with

little or no conscious reasoning. Yet, these rapid processes are flexible, not fixed: new sentences are processed, new scenes are comprehended, important items are selected for attention even though they cannot be explicitly anticipated, novel sentences are formulated to express an idea, and appropriate actions are taken. I propose that CSTM plays an essential role in these processes. The working of CSTM is best revealed when two or more stimuli are presented together or in a rapid sequence, as in rapid serial visual presentation (RSVP), or when a rich stimulus is presented, such as a picture of a scene.

Unlike STM, CSTM is central to cognitive processing. Recognition of meaningful stimuli such as words or objects rapidly activates conceptual information and leads to the retrieval of additional relevant information from LTM. New links among concurrently active concepts are formed, guided by parsing mechanisms of language or scene perception and by higher-level knowledge. When these new links result in well-connected structures, the structures are likely to be consolidated into LTM. Information that is not incorporated into such structures is rapidly forgotten. This whole cycle—identification of stimuli, memory recruitment, structuring, consolidation, and forgetting of nonstructured material—may occur in less than 1 sec when viewing a pictured scene or reading a sentence. (Potter 1993, p. 156)

The proposal that CSTM is a memory system distinct from STM and LTM is based on evidence for high-level processes that occur within a second of the onset of a stimulus, processes that depend on at least brief retention of stimuli at a conceptual level, together with associations that these conceptual representations activate from LTM. Standard short-term or working memory, such as Baddeley's articulatory loop and visuospatial sketchpad (e.g., Baddeley 1986), focuses on memory systems that support cognitive processes that take place over several seconds or minutes. A memory system such as the articulatory loop is unsuited for conceptual processing that takes place within a second of the onset of a stream of stimuli: it takes too long to be set up, and it does not represent semantic and conceptual information directly. Instead, STM directly represents articulatory and phonological information (the articulatory loop system) or visuospatial properties (the visuospatial sketchpad): these representations must be reinterpreted conceptually before further meaning-based processing can occur.

That approach neglects the evidence (some of it reviewed in the present chapter and elsewhere in this volume) that stimuli in almost any cognitive task rapidly activate a large amount of potentially pertinent

information, followed by rapid selection and then decay or deactivation of the rest. That can happen an order of magnitude faster than the setting up of a standard, rehearsable STM representation, permitting the seemingly effortless processing of experience that is typical of cognition. Of course, not all cognitive processing is effortless: our ability to engage in slower, more effortful reasoning, recollection, and planning may well draw on conventional short-term memory representations. I return to this point later.

The proposed architecture is similar in spirit to such contemporary models of processing as Anderson's ACT* (1983; see also Anderson 1993), Kintsch's (1988) construction-integration model of discourse comprehension, Ericsson and Kintsch's (1995) theory of long-term working memory (LT-WM), and Just and Carpenter's model of reading comprehension (1992). While these models differ from each other in many respects, all assume some form of processing that relies on activation or memory buffers other than standard STM. Thus, the idea of immediacy of processing (with a brief memory buffer) is not new, although it tends to be neglected in favor of the slower processes of conventional working memory.

Evidence for CSTM—An Overview

The CSTM hypothesis consists of several interrelated claims. First, presentation of a sequence of meaningful stimuli gives rapid access to semantic information about each stimulus, including its associations. Second, this information is used in various ways, depending on the viewer's current goal: if the viewer is trying to understand the whole sequence (e.g., a sentence), the information is used to discover or build a comprehensive structured representation, but if the viewer is trying to locate and identify a particular kind of information (as in target search), then only a subset of the information is selected. Third, whatever information has not been incorporated in such a structure, or selected as a target of interest, is highly likely to be forgotten, often before it enters awareness. A brief review of evidence for each of these assumptions of the model follows. A more detailed discussion of some of this work follows this overview.

1. *There is rapid access to semantic information about a stimulus.* As recently as the early 1970s it was still unclear whether semantic information was retrieved as part of STM or only subsequent to phonemic encoding (e.g., Shulman 1971), even when words to be remembered were each presented for 1 sec or longer. Since that time it has become evident that conceptual information about a

stimulus such as a word or a picture is available within 100–300 msec. Among the experimental paradigms that have shown such rapid availability are semantic priming (Neely 1991), including masked priming and so-called fast priming (Sereno & Rayner 1992); eye tracking of reading (Rayner 1983) and of scene perception (Loftus 1983); measurement of event-related potentials during reading (Kutas & Hillyard 1980); and target detection in rapidly presented sequences of pictures (Potter 1976), words (Lawrence 1971b), and letters and digits (Chun & Potter 1995; Sperling, Budiansky, Spivak, & Johnson 1971). These and other studies, some of which are discussed in more detail later in this chapter, show that semantic or conceptual factors have an effect on performance within a few hundred msec of the onset of the critical stimulus.

2. *This activated conceptual information can be used to discover or build a structured representation of the information, or to select certain stimuli at the expense of others.* A major source of evidence for this claim comes from rapid serial visual presentation (RSVP; Forster 1970) of words of a sentence, compared with scrambled sentences or lists of unrelated words. Studies by Forster (1970) and Potter (1984a; 1993; Potter, Kroll, & Harris 1980; Potter, Kroll, Yachzel, Carpenter, & Sherman 1986) show that it is possible to process the structure in a sentence and hence to recall it subsequently, at a rate such as 12 words/sec. In contrast, when short lists of unrelated words are presented at that rate, only 2 or 3 words can be recalled (see also Lawrence 1971a). For sentences, not only the syntactic structure but also the meaning and plausibility of the sentence are recovered as the sentence is processed (Potter et al. 1986). Because almost all sentences one normally encounters (and all the sentences in these experiments) include new combinations of ideas, structure-building is not simply a matter of locating a previously encountered pattern in long-term memory: it involves the instantiation of a new relationship among existing concepts. Structure-building presumably takes advantage of as much old structure as possible, using any preexisting associations and chunks of information to bind elements (such as individual words in a list) together.

Selective processing based on rapid access to information about the identity and meaning of stimuli is shown in serial search for targets that are specified by category (e.g., “an animal,” “a letter”), as in recent work on the attentional blink (e.g., Chun & Potter 1995; Raymond, Shapiro, & Arnell 1992; Potter, Chun, Banks, & Muckenhoupt 1998) and in older picture search (e.g., Intraub 1981;

Potter 1975, 1976) and word search (Lawrence 1971b) studies. To detect a target defined by its category, the target must first be identified. The finding in all these experiments that targets can be detected at rates of 8–10 items/sec or higher shows that categorical information about a stimulus is activated and then selected extremely rapidly.

3. *There is rapid loss of information that does not become structured or that is not selected for further processing.* The CSTM hypothesis is not only that conceptual information is activated rapidly, but also that the initial activation is highly unstable, such that the information is deactivated or forgotten within a few hundred msec if it is not incorporated into a structure (or selected for further processing). (Note the similarity between this assumption and that of rapid decay of iconic memory; however, precategorical iconic memory is clearly distinct from CSTM. Theories such as that of M. Coltheart [e.g., 1983] include a postcategorical stage in iconic memory, which may be identified with CSTM.) The assumption is that as a structure is built—for example, as a sentence is being parsed and interpreted—the resulting interpretation can be held in memory and ultimately stabilized or consolidated in longer-term memory as a unit, whereas only a small part of an unstructured sequence such as a string of unrelated words can be consolidated in the same time period.

I use the term “consolidation” descriptively; the nature of the process that results in a more stable representation, in either STM or LTM, is not yet known. Consolidation of information from CSTM into a more stable representation such as STM appears to operate serially on single items, chunks, or connected structures, and to require time, as shown in studies of picture memory and studies of the attentional blink already cited. The importance of structuring and of study time in converting a short-term memory into a long-term memory is well recognized; as a rule of thumb, it has been suggested that it takes 5 sec per item to establish a long-term memory. However, a single item in CSTM appears to take an order of magnitude less time to become stabilized in STM: 500 msec or less. When a sequence of items such as the words in a sentence can be structured rapidly in CSTM, stabilization in reportable memory seems to occur simply as a consequence of structuring, just as comprehension of a conversation or story results in a long-term memory representation. Evidence for rapid forgetting of material that is not well-structured is discussed below.

Rapid Processing Followed by Rapid Forgetting: Case Studies Using RSVP

The original motivation for presenting static stimuli such as still photographs in a continuous sequence was to simulate normal visual perception, in which the eye fixates briefly on a succession of points and thus processes a continuous sequence of snapshots (Potter & Levy 1969). Similarly, in reading or in listening to speech there is a steady flow of new information: events do not occur in single, isolated trials. In these circumstances rapid conceptual activation is often followed by rapid forgetting of some of the material. In this section I discuss a number of studies using RSVP as a tool to investigate rapid comprehension and rapid forgetting in CSTM.

Selective Search and the Attentional Blink

In brief, the attentional blink (AB) is a phenomenon that occurs in RSVP search tasks in which two targets are presented among distractors. When the rate of presentation is high but still compatible with accurate report of a single target (e.g., a presentation rate of 10/sec when detecting a letter among digit distractors), two targets are also likely to be reported accurately—except when the second target appears within 200–500 msec of the onset of the first target. This interval during which second-target detection drops dramatically was termed an attentional blink by Raymond et al. (1992). Shapiro and Luck's chapter in this volume provides a comprehensive review of the literature on AB, so in the following I focus on Chun's and my research as it pertains to CSTM (Chun 1997a, 1997b; Chun, Bromberg, & Potter 1994; Chun & Potter 1995; Potter, Chun, et al. 1998).

The attentional blink is relevant to CSTM because it provides evidence for rapid access to categorical information about rapidly presented items and at the same time shows that selective processing of specified targets has a cost. AB experiments suggest that there is a difference in time course between two stages of processing, a first stage that results in identification of a stimulus (CSTM) and a second stage required to consolidate that information in a reportable form, when the task is to pick out targets from among distractors (Chun & Potter 1995).

Consider a task in which targets are any letter of the alphabet, presented in an RSVP sequence of digit distractors. Presumably a target letter must be identified in order to be classified as a target (see Sperling et al. 1971). At rates as high as 11 items/sec the first letter target (T1) is detected quite accurately, consistent with evidence that a letter can be identified in less than 100 msec. We term this initial identification

Stage 1 processing, which constitutes activation of a conceptual but short-lasting representation, a CSTM representation.

A second target letter (T2) that arrives soon after the first one is likely to be missed, suggesting that a selected target requires additional processing beyond identification: Stage 2 processing. It is Stage 2 processing of T1 that interferes with the processing of T2. Stage 2 processing is necessary to consolidate a selected item into some form of short-term memory that is more stable than CSTM. We further hypothesized that the items following the first target (T1) continue to be processed successfully in Stage 1 and remain for a short time in CSTM; the problem is that as long as Stage 2 is tied up with T1, a second target may be identified but must wait, and thus may be lost from CSTM before Stage 2 is available. When this happens, T2 is missed, producing an attentional blink. Although the duration of the AB varies, it is marked at 200 msec after the onset of the first target and diminishes over the next 300 msec; it is usually gone by 500 msec.

Recently, several findings have supported the Chun–Potter hypothesis that T2 does receive Stage 1 processing, even if it is subsequently unavailable for report because it is “blinked.” Luck, Vogel, and Shapiro (1996) found that an unreported T2 word (falling in the AB interval) nonetheless resulted in a significant N400 cortical evoked potential based on its meaning. Maki, Frigen, and Paulson (1997) found that an unattended distractor falling in the AB interval could semantically prime a second target also falling in that interval, and the size of the priming effect was as great as that between the prime-target pairs outside the blink interval. They interpreted this result as consistent with the claim of the Chun–Potter two-stage model that all (or most) items are processed to a conceptual level, whether or not they can be reported. Shapiro, Driver, Ward, and Sorenson (1997) also obtained evidence for semantic activation of targets that appear in the attentional blink window and fail to be reported.

In Chun and Potter's model, Stage 1 identification of the first target initiates the Stage 2 process of attention and consolidation. However, the attentional process that selects the target for second-stage processing is temporally inexact, so that frequently the target and the following item are both passed to Stage 2 (Raymond et al. 1992 proposed a similar hypothesis). When a distractor that is confusable with a target (e.g., a digit distractor with a letter target) is processed together with the target in Stage 2, sorting out and consolidating the target is slower than it would be with no following distractor, or with a dissimilar distractor such as an asterisk (Chun & Potter 1995; see also Maki, Couture, Frigen, & Lien 1997). When the immediately following item

is the second target, then the inexact attentional process is likely to pick up both T1 and T2 for processing in Stage 2, and both targets are likely to be successfully processed, producing the lag 1 sparing (the check mark profile as a function of lag) that is frequently observed in AB experiments. But if even a single distractor intervenes between T1 and T2, then T2 waits: thus, the AB effect is reduced or absent at lag 1 and maximal at lag 2, when the wait for Stage 2 processing is longest on average.

Is There a Cross-Modal Attentional Blink? Potter, Chun, et al. (1998; see also Potter, Chun, & Muckenhoupt 1995) reported a series of experiments in which compressed speech was used to create an auditory equivalent of RSVP. Subjects listened to a rapid series of spoken digit distractors and two letter targets. In this condition we found some deficit for the second letter, but critically there was no effect of the lag between the two spoken letters, and thus no indication of a transient auditory "blink." Similarly, when the sequence started out in one modality and then switched to the other modality, with one target in each modality, there was again no evidence for a blink. However, Arnell and Jolicoeur (1995, 1997) did find evidence for a lag-dependent deficit in auditory and cross-modal conditions that they considered to be an attentional blink. To attempt to resolve this difference in results, Potter, Chun, et al. (1998) differentiated the visual attentional blink from another attentional deficit in serial search tasks that we believe to be the consequence of task-switching.

Costs of switching between one task and another are standardly studied by comparing performance on the first trial of the new task with performance on the second or subsequent trials. In many conditions, the first trial after a task switch is slower or less accurate than subsequent trials, even though the task is highly practiced and the participant is informed that a switch is about to occur (Allport, Styles, & Hsieh 1994; Meiran 1996; Monsell 1996; Rogers & Monsell 1995). These experiments (with one exception) used single trials in which the subject responded to each stimulus before seeing the next one.

Consider, however, what one might expect in the case of serial search, if the second of two targets requires a switch in perceptual set. Potter, Chun, et al. (1998) noted that in some AB studies (e.g., Chun & Potter 1995) the two targets are defined in the same way (two letters among digits), whereas in other studies the two targets are defined differently. In a typical procedure of the latter type, the first target is a white letter among black letter distractors and the second target is a black letter X among other letters: the X is present on a random half of the trials (Raymond et al. 1992). The viewer has to report the identity of the

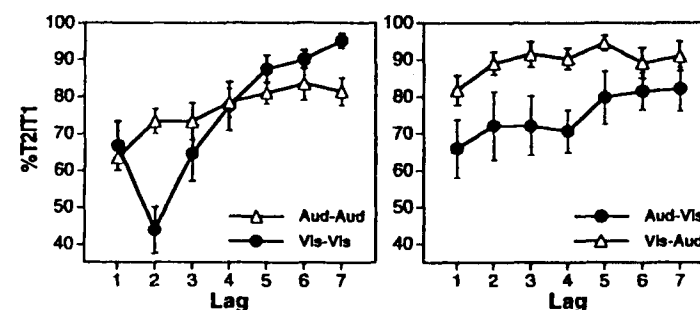


Figure 2.1

Correct report of T2 (given correct report of T1) as a function of lag, with 2 simultaneous sequences of visual and auditory items (SOA = 120 msec) and instructions to identify T1 and detect "X" (T2) in specified modalities, for the groups with 2 visual or 2 auditory targets (panel at left), and the groups with 1 visual and 1 auditory target (panel at right). Only trials on which an "X" was presented are included. Vertical bars show standard error of the mean. (Potter, Chun, et al. 1998, experiment 4.)

white letter (T1) and then report the presence or absence of T2 (the X). Thus, there is a switch from a set to pick out and identify the white letter, ignoring black letters, to a set to make a presence-absence decision as to a black X. Even though subjects do not make an overt response until after both targets have been presented, such a switch in perceptual and (covert) response set could take time, during which T2 might be missed. The covert task switch would produce a lag-dependent deficit that is maximal at lag 1, the immediately following stimulus (in contrast to the attentional blink, which shows sparing at lag 1).

In most of Arnell and Jolicoeur's experiments showing cross-modal and auditory attentional deficits (1995, 1997), they used different tasks for T1 and T2. Potter, Chun, et al. (1998) hypothesized that the deficits Arnell and Jolicoeur observed in the cross-modal and auditory conditions were actually task-switch deficits; one cue was that in the cross-modal conditions the largest deficit was at lag 1. We replicated their findings when we adopted their task-switch procedure (figure 2.1), but when we used the same letter-detection task (or digit-detection task) for both T1 and T2, we attenuated or eliminated the lag-dependent deficit—except for the all-visual condition (figure 2.2). Significantly, in the task-switch condition the greatest deficit was at lag 1—except for the all-visual condition, in which lag 1 reflected a combination of the visual AB pattern of sparing and the task-switch deficit. The lag 1 benefit reflects the properties of visual stimuli, which are unstable in visual STM or CSTM and must be stabilized by further processing in Stage 2. In contrast, auditory stimuli appear to enter an auditory buffer

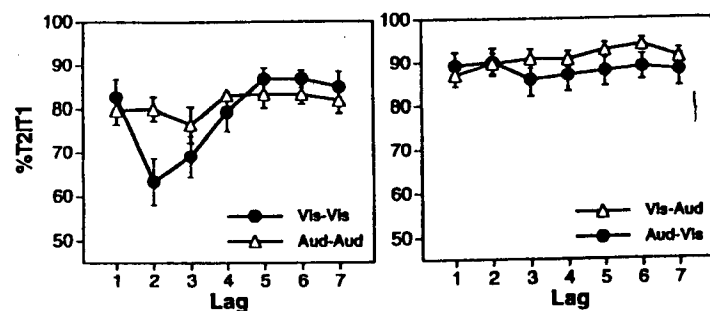


Figure 2.2
Correct report of T2 (given correct report of T1) as a function of lag, with 2 simultaneous sequences of visual and auditory items (SOA = 120 msec), when the task was to report 2 letters among digit distractors. The results are shown separately for each of the 4 conditions, VV, AA, VA, and AV. Vertical bars show standard error of the mean. (Potter, Chun, et al. 1998, experiment 5.)

that has a longer time course, reducing or eliminating lag effects as long as the task is consistent from T1 to T2. Task-switching costs, however, have to do with central set and presumably are not modality-dependent, which is why they show up in auditory and cross-modal conditions. When a task-switching procedure is combined with all-visual presentations, then both the standard AB effects and task-switching effects are observed; lag 1 benefits may or may not be found.

Summary: CSTM and AB Studies of the visual attentional blink demonstrate a dissociation between an early stage of processing sufficient to identify letters or words presented at a rate of about 10/sec, and a subsequent stage of variable duration (up to about 400 msec) required to stabilize a selected item in reportable STM. The attentional blink thus provides evidence for the central claims of CSTM.

Repetition Blindness and Token Failure

Helene Intraub was the first person to discover the repetition blindness phenomenon, when she was investigating conceptual short-term memory for pictures in the late 1970s as a postdoctoral fellow in my lab. She noted that the repeated picture she had placed in a sequence seemed to disappear when she ran the film, even though there were other pictures intervening between the two presentations: she had to check the film slowly to confirm that the picture had indeed appeared twice. She called me into the lab to observe this surprising illusion. Kanwisher (1987) investigated this illusion with sequences of words, in lists and sentences; she dubbed the effect "repetition blindness"

(RB). She confirmed that the second of two identical visual stimuli is often not noticed when it appears in an RSVP sequence shortly after the first stimulus. What is remarkable is that items that intervene between the first instance (C1) and the second one (C2) may be perceived and reported accurately. Like the attentional blink, RB shows a lag effect, although Chun (1997a) has found that, unlike visual AB, RB is maximal at lag 1.¹ Using the AB target-search procedure, Chun showed that AB and RB are doubly dissociable: RB is found for second identical targets (or for targets identical except for letter case) even when the distractors are keyboard symbols and there is no AB, and RB turns into a repetition benefit when the targets (letters among digits) are redundantly signaled by being colored rather than black, whereas AB is substantial in this condition.

Kanwisher (1987) proposed that repetition blindness occurs when the viewer fails to set up a token (or object file; Kahneman & Treisman 1984) of the second stimulus. Tokens are contrasted with types: types are the long-term representations of types or categories of objects, including words, that are used in the recognition of objects. When an object or word is viewed, recognition requires not only that the appropriate type representation be activated, but also that a representation or token of the object's presence be set up in this particular episode (see Kanwisher's and Bavelier's chapters in this volume for further discussion of this theory). Without a token representation, the occurrence of an item is unreportable. The loss of reportable information about the occurrence of the second of two identical items (RB) is due to a bias in the visual system against the immediate retokenizing of the same type: instead, the second occurrence, which does activate the type representation, is taken to be part of the first occurrence, which has already been tokenized.

One can ask whether CSTM consists of type activation or new tokens or both. CSTM is a form of representation in which old (recognizable) items are activated to form new structures, so it is clear that existing types (and their associations) must be activated in CSTM. Insofar as structuring occurs in CSTM processing, the new structures must be represented by establishing tokens of the relevant types and their relationships. For example, to process even a simple novel sentence successfully requires the activation of one's knowledge of the word types, of associations among the concepts constrained by parsing mechanisms, and a token representation of the resulting meaningful structure. The finding that syntactic and pragmatic constraints have little or no effect on repetition blindness for the second occurrence of a word in a sentence (Kanwisher 1987; Kanwisher & Potter 1990) suggests that failure to create a token for the second occurrence of a word prevents it

from entering into the structuring process, so that the sentence "Nancy spilled the ink and there was ink all over" is processed and recalled as the ungrammatical string "Nancy spilled the ink and there was all over." For sentence processing, a failure to include a token of the second "ink" apparently makes it invisible to the parser.

Thus, although structuring in CSTM makes use of types and their connections as represented in LTM, it seems likely that copies or tokens of these activated types are what enter into the structuring process in CSTM, at least in the case of syntactic processing. Repetition blindness would prevent an item from participating in CSTM. On the other hand, Bavelier (in the present volume; Bavelier 1994; Bavelier & Potter 1992) has argued that some forms of repetition blindness, particularly those between nonidentical visual stimuli that share only an identical phonological representation (e.g., a picture of the sun and the written word *son*), arise only after an initial token has been opened for the second item; as more information (e.g., phonological information) becomes activated, the opened token will become subject to RB if the added information leads to sufficiently similar representations of C1 and C2.

That is, Bavelier proposes that tokenization is not an all-or-nothing process, but occurs over a period of tens or hundreds of milliseconds as more information about a type is accrued, either stabilizing the token or, in RB, making it similar enough to an earlier token to cause the two to merge. This view implies that such items are represented at least partially in CSTM, albeit briefly. Such a fleeting representation may account for the preservation of semantic priming from a repeated homophone in lexical decision (e.g., none . . . nun PRIEST) despite evidence for RB for *nun* in recall (Coltheart, this volume). Bavelier notes that RB effects tend to be considerably weaker between visually dissimilar items than between highly similar or identical visual items, suggesting that much of the time both items are at least briefly available in CSTM.

In understanding the basis for RB it is important to note that visual similarity and phonological similarity of the names of visual stimuli can both produce RB, but there is little or no evidence for conceptually based repetition blindness. Synonyms such as "couch" and "sofa" do not produce RB (Altarriba & Soltano 1996; Kanwisher & Potter 1990; but see MacKay & Miller 1994; and MacKay, Abrams, Pedroza, & Miller 1996). This suggests that the tokenizing process is concerned with the representation of visual entities, not their meanings. A further question is whether an analog of RB is found with auditory stimuli: repetition deafness (RD). While Miller and MacKay (1994, 1996) did report evidence for RD in lists of words (but not sentences), Downing, Kanwisher, and Potter (1998) failed to find either RD or a cross-modal repetition

deficit; if anything, they obtained a repetition benefit (positive priming) in the auditory case. As Potter, Chun, et al. (1998) speculated with respect to the absence of an auditory or cross-modal attentional blink, it seems likely that the auditory system has a robust mechanism for representing rapid sequences of sounds, a mechanism that would buffer rapidly changing information more effectively than in the visual system.

Summary: RB The loss of a stimulus when it is identical or similar to a stimulus seen within the last 0.5 sec points to the vulnerability of rapidly presented stimuli to forgetting. The very structuring process that stabilizes associated items in CSTM appears to collapse two events into one when they arrive in quick succession and are categorized visually, orthographically, or phonologically under the same heading—even when they are in fact distinct not only temporally but also in terms of their visual properties (uppercase versus lowercase letters, homophones, pictures and words, etc.).

Understanding RSVP Sentences

Given the marked problems in encoding just two stimuli in the target-search tasks used to investigate the attentional blink, it is striking that one can read an RSVP sentence at the same rate (10 words/sec), and both understand and recall it (Potter 1983, 1984a; Potter et al. 1986). Although part of the difference between search tasks and sentences is due to the special processing demands when selecting a target from among distractors, another major factor appears to be the difficulty of retaining unrelated items even briefly. In this section I review some of the evidence for this claim, which is central to the CSTM hypothesis.

Differences Between Lists and Sentences The memory span for words is 5 or 6, when the presentation rate is consistently 1/sec. But when lists of 2, 3, 4, 5, or 6 nouns were presented at higher rates using RSVP, I found that immediate recall declined to a mean of 2.6 words for 5-word lists (2.4 for six-word lists) at the rate of 12 words/sec, as shown in figure 2.3 (Potter 1982, 1993). This was evidently not because participants could not recognize the words at that rate, because a list of just 2 words (followed by a mask) was recalled almost perfectly at 12/sec: instead, some additional process was necessary to stabilize the words in short-term memory. At the rate of 10/sec, about 3 words were remembered (for lists of 3 to 6 words), at 3/sec, about 4 words (for lists of 4 to 6 words). In another study I found that the presentation of two related words on a 5-word RSVP list (separated by another word) resulted in improved recall for both words, suggesting that both words were activated to a level at which an association could be

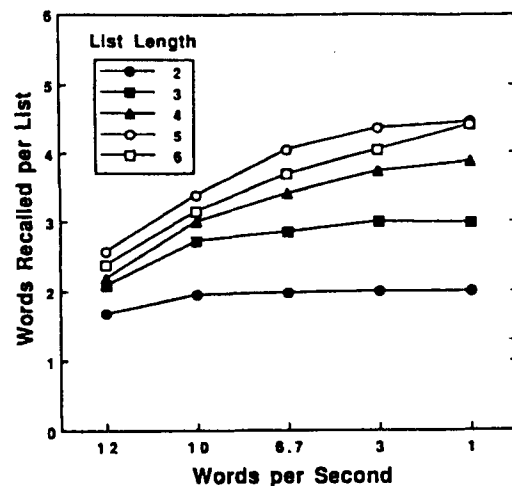


Figure 2.3
Immediate recall of RSVP lists of 2, 3, 4, 5, and 6 nouns presented at rates between 1 and 12 words/sec (Potter 1982). A mask followed each list. (Adapted from Potter 1993, figure 2.)

retrieved. This hinted at the sort of process that might stabilize or structure information in CSTM.

In contrast to lists, 14-word sentences presented at rates up to at least 12 words/sec can be recalled quite accurately (see Potter 1984a; Potter et al 1986), at least if they are not syntactically complex and if they convey straightforward ideas. The findings with lists versus sentences strongly support the CSTM assumption that each word can be identified and understood with an 83–100 msec exposure, even when it is part of a continuing series of words. The results also support the second assumption that representations of the words remain activated long enough to allow them to be bound into whatever syntactic and conceptual structures are being built on the fly. When, as with a list of unrelated words, there is no ready structure to hand, all but 2 or 3 of the words are lost.

How Are RSVP Sentences Remembered? The Regeneration Hypothesis Before addressing the question of how rapidly presented sentences are retained, one should address the prior question of why sentences heard or read at *normal* rates are easy to repeat immediately, even when they are two or three times as long as the list that can be repeated accurately. The difference in capacity between lists and sentences is thought to be due to some form of chunking, although it has also been assumed that sentences can be stored in some verbatim form temporarily (see the

review by Von Eckardt & Potter 1985). Potter and Lombardi (1990) proposed a different hypothesis: immediate recall of a sentence (like longer-term recall) is based on a conceptual or propositional representation of the sentence. The recaller regenerates the sentence, using normal speech production processes to express the propositional structure. We proposed that recently activated words were likely to be selected to express the structure. In consequence, the recalled sentence is normally verbatim, but not because there is a sequential verbatim representation of the words (e.g., a phonological representation) that is simply parroted.

To test this hypothesis, Potter and Lombardi (1990) presented distractor words in a secondary task immediately before or after the to-be-recalled sentence, and on some trials one of the words was a good substitute for a word in the sentence (such as “castle” for “palace”). As we predicted, that word was frequently intruded in recall, as long as the rest of the sentence was consistent with the substitution. Thus, recall was guided by a conceptual representation, not by a special verbatim representation.

Further studies (Lombardi & Potter 1992; Potter & Lombardi 1998) indicated that syntactic priming from having processed the sentence plays a role in the syntactic accuracy of immediate recall of sentences. Syntactic priming (e.g., Bock 1986) is a temporary facilitation in the production of a recently processed syntactic structure, as distinguished from direct memory for the syntactic structure of the prime sentence. Among the reasons that sentences are no longer recalled verbatim after one intervening sentence (e.g., Sachs 1967, 1974) are that the conceptual structure is now more complex (if the sentences are related), the relevant words are no longer as activated, and syntactic priming may have decayed or been interfered with by the intervening sentence.

The Potter-Lombardi hypothesis that sentences are regenerated rather than “recalled verbatim” is consistent with the CSTM claim that propositional structures are built rapidly, as a sentence is read or heard. More directly relevant is one of the Potter-Lombardi (1990) experiments in which the sentences were presented at a rate of 10 words/sec, rather than the moderate 5 words/sec of their other experiments: the intrusion results were essentially the same, showing that the relevant conceptual processing had also occurred at the higher rate.

Reading RSVP Paragraphs: More Evidence for Immediate Use of Structure A single RSVP sentence apparently is easy to comprehend and recall when presented as fast as 12 words/sec, so that recall is close to ceiling. Does that mean that longer-term retention of the sentence will be as good as if it had been presented more slowly? To answer that

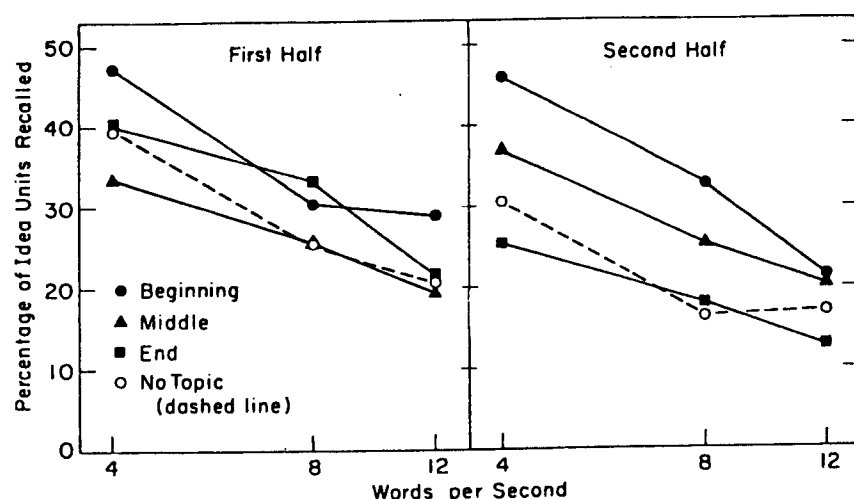


Figure 2.4
RSVP paragraphs presented at 3 rates; percentage of idea units recalled in each half of the paragraph, as a function of the position of the topic sentence in the paragraph. (Adapted from Potter et al. 1980, figure 20.2.)

question, we (Potter et al. 1980) presented RSVP paragraphs of about 100 words at three rates: 4, 8, and 12 words/sec, with the equivalent of a two-word pause between sentences (the net rates averaged 3.3, 6.7, and 10 words/sec). Immediately after presentation, participants wrote down the paragraph as accurately as possible. To allow us to evaluate both single-word perception and use of discourse-level information, we used paragraphs that appeared to be ambiguous and poorly integrated unless the reader knew the topic (see Bransford & Johnson 1972; Dooling & Lachman 1971). We included a sentence that mentioned a one-word topic (e.g., "pizza") at the beginning, the middle, or the end of the paragraph, or we omitted the topic. Our predictions were that the topic word would be recalled, that any part of the paragraph that followed the topic would be recalled more accurately than any part that preceded it (so that having the topic at the end of the paragraph would be no better than omitting it entirely), and that both predictions would be true at all rates of presentation. We also expected that the higher the rate, the less the recall.

Whether we scored only verbatim recall or used a more liberal score of idea units recalled, recall was improved after the topic was presented (but not before), at all three rates of presentation: therefore, even at the highest rate the discourse topic could be used to structure the paragraph (figure 2.4). This suggests that the discourse topic, once it

became evident, remained active as a source of structure as the rest of the paragraph was read. (The topic word was perceived and recalled by more than 80% of the subjects regardless of rate or condition). At the same time, there was a marked main effect of rate: recall declined as rate of presentation increased, from 37% of the idea units at 4 words/sec to 26% at 8 words/sec to 20% at 12 words/sec, averaging over all topic conditions. Clearly, even though there was internal evidence that discourse-level structuring was occurring at all rates of presentation, some process of consolidation was beginning to fail as rate increased.

Putting the paragraph results together with those for lists and single sentences, we see that structuring can occur rapidly, and more structure results in better memory (comparing lists with sentences, or comparing a string of seemingly unrelated sentences with sentences structured by having a topic). Nonetheless, rapid conceptual processing is not sufficient for accurate retention if there is no additional time for consolidation: the gist may survive, but details will be lost in immediate recall, just as they are in longer-term memory.

Mechanisms of Structuring in RSVP Sentence Processing Although I have repeatedly invoked the idea that there is rapid structuring of information represented in CSTM, I have had little to say about just how this structuring occurs. In the case of sentences, it is evident that parsing and conceptual interpretation must occur virtually word by word, because any substantial delay would outrun the persistence of unstructured material in CSTM (as one sees in the case of the attentional blink). Here I will briefly describe three studies that have investigated the process of selecting an appropriate interpretation of a given word in an RSVP sentence, a key process in comprehension, given the extent of lexical ambiguity in English and in most other languages.

THE INFLUENCE OF SENTENCE CONTEXT ON WORD AND NONWORD PERCEPTION In one study (Potter, Moryadas, Abrams, & Noel 1993) we took advantage of the propensity of RSVP readers to convert a nonword into an orthographically similar word. We presented nonwords such as *dack* that are one letter away from two other words (*deck*, *duck*), in sentences biased toward one or the other of these words or neutral between them, as in the following examples. Note that when we presented a real word in the biased sentences, it was always the mismatching word. Subjects recalled the sentence; they were told to report misspelled words or nonwords if they saw them.

Neutral: "The visitors noticed the *deck/duck/dack* by the house."

Biased: "The child fed the *deck/dack* at the pond."

"The sailor washed the *duck/dack* of that vessel."

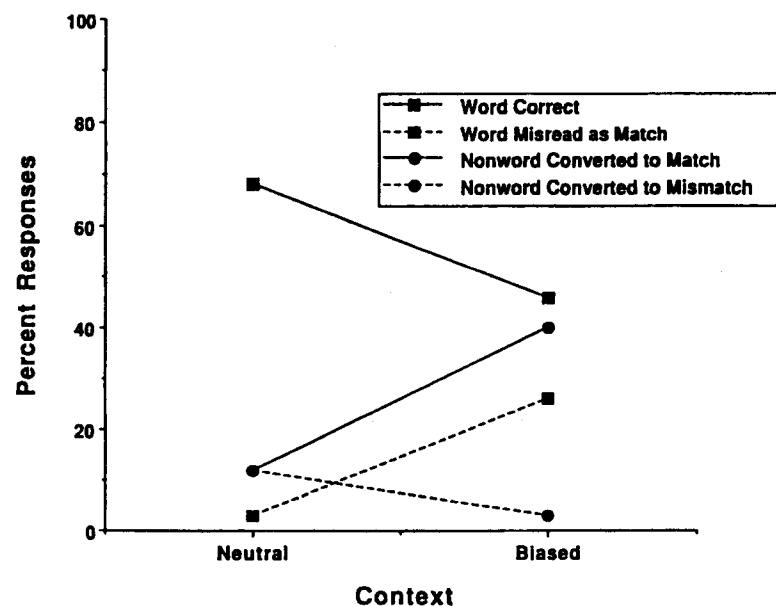


Figure 2.5

Percent report of the critical word (always the mismatch word) or nonword (similar to both the match and mismatch words) as the match or mismatch word. In the neutral condition the match and mismatch words were equally plausible; in the biased context only the match word was plausible. Subjects were instructed to report incongruous or misspelled words, if any. (Adapted from Potter et al. 1993, experiment 2.)

As figure 2.5 shows, we found that readers were much more likely to convert the nonword in the biased direction (40% of trials) than in the other direction (3% of trials). Similarly, when the inappropriate word was put in the biased sentence, misreadings increased dramatically and accurate reports dropped, although the incongruous word was still reported correctly on almost half the trials. Thus, context can bias word and nonword perception even when reading at 10 words/sec. More surprisingly, we found that even selective context that does not appear until as much as three words (300 msec) after the critical word or nonword can influence perception, suggesting that multiple word candidates (and their meanings) are activated as the nonword or word is perceived, and may remain active for at least 300 msec after the word or nonword has been masked by succeeding words. This supposition that multiple possible words and their meanings are briefly activated during word perception accords with the Swinney hypothesis (1979) that multiple meanings of ambiguous words are

Table 2.1

Percentages of double words (matching and nonmatching) recalled in each context condition (before and after) as part of the sentence.

Double Word	Context Condition	
	Before	After
Matching	75	65
Nonmatching	10	15

Before: Maggie carried the kitten in a ^{basket}***** to her house.
pencil

After: Maggie used a ^{basket}***** to carry the kitten.
pencil

Adapted from Potter et al. 1998, experiment 1.

briefly activated: both results are consistent with the CSTM view. In the present study and in the case of ambiguous words, the process of activation and selection appears to occur unconsciously for the most part, an issue considered in a later section.

DOUBLE-WORD SELECTION In another study (Potter, Stiefbold, & Moryadas 1998) we presented two orthographically distinct words simultaneously (one above and one below the line) in the course of an RSVP sentence, instructing the participant to select the one that fit into the sentence and include it when immediately recalling the sentence. We regarded this as an overt analog of lexical ambiguity resolution. The sentence was presented at about 7.5 words/sec; the two-word ("double word") display, for 83 msec. As table 2.1 shows, sentence context had a massive influence on selection, both when the relevant context arrived before the double words and when it arrived later (up to 1 sec later, in one experiment), showing that readers could activate and maintain two distinct lexical possibilities. Subjects were asked to report the "other" word (the mismatching word) after they recalled the sentence, but most of the time they were unable to do so, showing that the unselected word was usually forgotten rapidly. Again, this illustrates the existence of fast and powerful processors that can build syntactically and pragmatically appropriate structures from briefly activated material, leaving unselected material to be rapidly forgotten.

LEXICAL DISAMBIGUATION Miyake, Just, and Carpenter (1994) carried out two experiments on self-paced reading of sentences with ambiguous words that were not disambiguated until 2–7 words after the ambiguous word. They found that readers with low or middling read-

ing spans were slowed down when the disambiguation was toward the subordinate meaning, especially with a delay of 7 words. (High-span readers had no problems in any of the conditions.) In an unpublished experiment we presented subjects with a similar set of sentences that included an ambiguous word, using RSVP at 10 words/sec; the task was to decide whether or not the sentence was plausible, after which we gave a recognition test of a subset of the words, including the ambiguous word. Our hypothesis was that sentences that eventually turned out to require the subordinate meaning of an ambiguous word would sometimes be judged to be implausible, implying that only the dominant reading had been retrieved. Unambiguously implausible and plausible sentences were intermixed with the ambiguous sentences.

Subjects were more likely to judge a plausible sentence to have been implausible when a subordinate meaning of the ambiguous word was required (27% versus 11% errors), when the disambiguating information appeared after a greater delay (23% versus 16% errors), and especially when there were both a subordinate meaning and late disambiguation (32% errors, versus 9% for the dominant/early condition). A mistaken judgment that the sentence was implausible suggests that on those trials only one meaning, the wrong one, was still available at the point of disambiguation. Interestingly, the ambiguous word itself was almost always correctly spotted on a recognition test of a subset of words from the sentence, even when the sentence had mistakenly been judged implausible. The results suggest that although multiple meanings of a word are indeed briefly activated (in CSTM), the less frequent meaning will sometimes fall below threshold within a second, when sentences are presented rapidly.

Understanding Pictures and Scenes

In the previous sections I have focused on CSTM as it is revealed in studies in which letters, digits, words, and sentences were presented in RSVP. In this section I review evidence that comprehension of pictures or scenes also involves rapid understanding followed by rapid forgetting.

In an early study (Potter & Levy 1969) 128 color photographs of a wide variety of scenes—close-ups of objects and people, indoor and outdoor scenes—were presented sequentially at rates between about 1 every 2 sec and 9/sec, followed by a recognition test of the presented pictures mixed with an equal number of new pictures. No picture was ever shown more than once. The rationale for presenting sequences of still pictures was that that is the way we normally take in visual information: by successive fixations (separated by brief saccades). With normal viewing, the stimulus onset asynchrony (SOA) between one

fixation and the next ranges from about 100 to 500 msec, averaging about 300 msec; longer fixations occur when the viewer focuses on a fine detail or a difficult-to-see stimulus.

In our studies we bracketed the normal range of fixation durations, concentrating on rates between 125 and 333 msec.² The main finding of our initial study (Potter & Levy 1969) was that the 16 pictures in a sequence were easily recognized in the test that followed the sequence if they had been presented for 1 or 2 sec each, but with shorter presentations, recognition memory declined, reaching almost chance at an exposure duration of 125 msec. These same pictures were easy to remember if they were presented singly for 90 or 120 msec, followed by a visual mask (Potter 1976): the problem seemed not to be that an exposure of 125 msec is too short to comprehend the picture, but rather that the presentation of the following to-be-attended picture cut off further processing in a way that the visual mask did not. I concluded that visual masking occurs with short SOAs (under 100 msec), whereas conceptual masking occurs with SOAs up to 500 msec or more (Potter 1976; see also Intraub this volume, 1980, 1981; Loftus & Ginn 1984; Loftus, Hanna, & Lester 1988).

In the next set of experiments (Potter 1975, 1976) I asked viewers of the picture sequences to detect a picture described by a brief title that was not explicit about pictorial details (e.g., "a boat," "a picnic on the beach," "two people drinking"). As shown in figure 2.6, the detection results were very different from recognition memory results with the same set of pictures: detection was above 60% at 125 msec per picture and above 80% at 250 msec per picture, when recognition memory (corrected for false yeses) was about 12% and 30%, respectively. Thus there was a strong indication that a viewer can comprehend a scene in 100–200 msec but cannot retain it without additional time (a median time of about 400 msec) for processing or consolidation, during which time it is vulnerable to conceptual masking from the next picture. This result is a prime example of CSTM, in a case in which there is no opportunity for linking the unrelated pictures into some kind of structure, so that most of the pictures in a sequence are simply forgotten at the higher rates.

Further studies by Intraub (1980, 1981, 1984; and see her chapter in the current volume) have provided more controlled tests of the disparity between visual search and later memory, and have examined the important role of attention. Intraub (1984) found that deliberately attending to the briefer of alternating brief and longer-duration pictures increased the probability of remembering the briefer pictures while decreasing the probability of remembering the longer pictures. This attentional trade-off is reminiscent of the attentional blink effect.

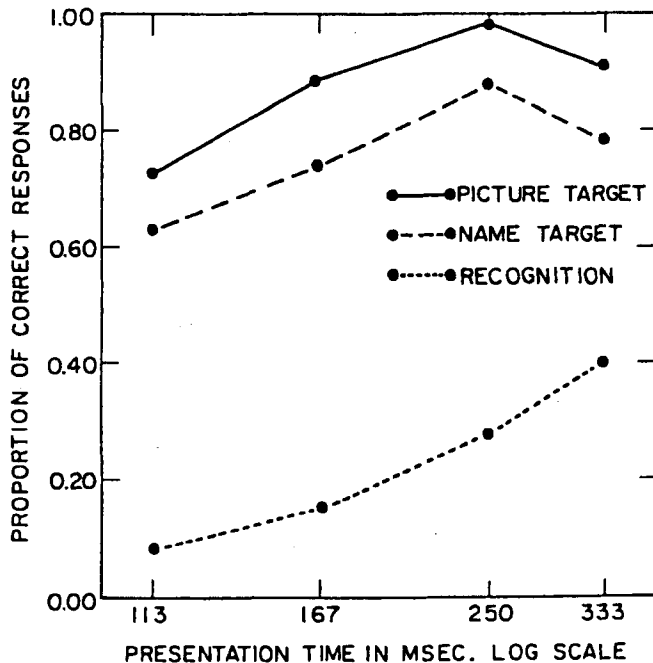


Figure 2.6
Proportion of correct detection responses to a target picture in a rapid sequence of pictures, when the target was shown in advance or was described by a brief title; the proportion of correct recognition responses (corrected for false yeses) is also shown. (From Potter 1976, experiment 1.)

Another approach to the question of picture comprehension versus picture memory is that of McConkie and Currie (1996) and Rensink, O'Regan, and Clark (1997), as well as the work of Wolfe (see his chapter in the present volume). McConkie and Currie, Rensink et al. have shown that our ability to recognize changes in a picture from one glimpse to the next (such as a change in size of the picture, or the addition or subtraction of one person from a group, a change in the color of a piece of clothing, a shift in the position of an item) is surprisingly poor; it is mainly a change in the object we are currently attending to (or planning a saccade to) that we notice. In these studies the change is made following a brief interruption to the scene, such as a blank interval of 80 msec or a saccade by the viewer, because a change made without such an interruption produces visual transients that attract attention. Some investigators (see O'Regan 1992) have concluded that we actually perceive much less in a scene than we subjectively suppose.

But another possibility, consistent with the CSTM hypothesis, is that we do perceive a great deal while we are actually viewing the scene, but only a subset of that information is still available once the next scene is presented: the new scene replaces the previous one, and little survives from the previous glimpse that can be compared with the present scene, other than the gist plus focally attended specifics.

Summary: Rapid Conceptual Processing Followed by Rapid Forgetting

In each of the experimental domains discussed—the attentional blink in selective search; repetition blindness; comprehension and retention of RSVP word lists, sentences, and paragraphs; studies of word perception and selection; and the experiments on picture perception and memory just reviewed—there is evidence for comprehension of the meaning or meanings of a stimulus early in processing (possibly before conscious awareness), followed by rapid forgetting unless conditions are favorable for retention. The two kinds of favorable conditions examined in these studies were selection for attention (e.g., T1 in the attentional blink procedure, and selection of a target picture from among rapidly presented pictures) and the availability of associations or meaningful relations between momentarily active items (as in sentence and paragraph comprehension and in word perception, selection, or disambiguation as a sentence is processed). The power of these two factors—selective attention that is defined by conceptual properties of the target, and the presence of potential conceptual structure—is felt early in processing, before conventional STM for the stimuli has been established, thereby justifying the claim that CSTM is separate from STM and working memory, as they are usually defined.

Further Questions About CSTM

As presented here and in Potter (1993), conceptual short-term memory is a functional construct that brings together diverse phenomena, all of which embody rapid conceptual activation of material that will be deactivated or forgotten almost immediately, but that remains active long enough for structure-building processes of perception, language, and thought to transform relevant material into a stable conceptual representation. CSTM is underspecified in many respects, and indeed no such construct could be fully specified until we know much more about the phenomena in question and about cognition more generally; the present volume moves us closer to that goal. In this section I raise some questions about CSTM and speculate about possible answers.

What Factors Determine Activation of Information in CSTM?

A distinction is often made between primary recognition or categorization of a stimulus and subsequent associations to that stimulus, but recognition and categorization themselves may involve associations, as was evident in misperception of a nonword like *dack* in the presence of a context word like *pond* versus *sailor* (Potter et al. 1993). A stimulus produces activation at many levels in the visual system and higher levels of processing, as incorporated in models such as those of McClelland and Rumelhart (1981) and Norris (1986) for word perception. This activation provides multiple possible interpretations of the stimulus at each level, requiring mechanisms for selecting the best fit among competing interpretations. Such mechanisms use semantic or pragmatic context associated with one or another interpretation to bias the outcome of the competition.

Psychologists have focused on word associations as an important source of memory activation and have used norms from the word association task as a measure of associative strength between words. Greater semantic priming between items with high than with low normative associations shows that word association norms capture something about the associative structure of the mind. However, the malleability of the word association task (which depends on the mental set of the subject, age, and the like) and the uncertainty about whether the norms measure association between lexical items, between concepts, or some combination of the two makes word association norms a questionable basis for a model of associative structure. Efforts to characterize and model human knowledge have had limited success: semantic networks, expert systems, schemata, prototypes, scripts, frames, and lexicographic approaches (to mention only a few examples) have each provided insights, but collectively they do not provide explicit constraints on what kinds of associations would be expected to result in activation of concepts in CSTM. All we can be sure of is that activation must be rapid, and hence only relatively direct associations are likely to be involved, at least in the first few hundred msec after presentation of a word (but see the discussion of LSA, below, for a different possibility).

What Are the Structuring Processes Within CSTM, and How Do They "Consolidate" Memory?

I assume that structuring in CSTM is not different in principle from the slower processes of comprehension that happen as we gradually understand a difficult text or an initially confusing picture, or solve a chess problem over a period of seconds and minutes. By definition, a difference between CSTM structuring and these slower processes is

the speed with which "solutions" are reached, and thus the relative absence of awareness that alternatives have been weighed and that many possibilities have been considered and rejected, at least implicitly. As in slower and more conscious problem-solving, a viewer's task or goal makes a major difference in what happens in CSTM, because one's intentions activate processing routines such as sentence-parsing, target specifications in search tasks, and the like. Thus the goal partially determines what enters CSTM and how structuring takes place.

The presence of many activated items at any moment, in CSTM, allows for compound cuing (e.g., McKoon & Ratcliff 1992)—the convergence of two or more weak associations on an item. The power of converging cues, familiar to any crossword puzzle fan, is likely to be central to structure-building in CSTM. A recent and radical proposal for the acquisition of knowledge (Landauer & Dumais 1997), latent semantic analysis (LSA), provides a suggestive model for how structure may be extracted from haphazard material. LSA's focus is on the slow buildup of "knowledge" of word meanings through massive exposure to texts, simply by analyzing the co-occurrence of words in paragraphs or other small units of text. The word co-occurrence matrix is subjected to an analysis similar to principal components analysis or factor analysis, which extracts *N* dimensions (factors or components) that capture the greatest variance in the matrix—300 dimensions were optimal in Landauer and Dumais's tests of LSA. A new text sample is interpreted by projecting it onto this 300-D space, a process that is fast. Although LSA is concerned with acquisition and only secondarily with comprehension, some such procedure may be involved in the rapid comprehension characteristic of CSTM (see also McKoon & Ratcliff 1998). However, there is no syntactic parser in LSA, and it is clear from RSVP research that we do parse rapidly presented sentences (see Potter, Stiefbold, & Moryadas 1998, experiment 4, for a recent example); thus, the LSA approach is at best a partial explanation of processing in CSTM.

Memory consolidation requires time, so that if the to-be-remembered material is presented sufficiently slowly, even arbitrary lists of items can be retained. But when information is presented more rapidly (as in RSVP), the more interconnected or structured or "chunked" the information to be remembered, the more time will be available to consolidate each chunk or unified structure—assuming that a unified structure can be consolidated in about the same time that an unconnected item can be. This is a version of Miller's original chunking hypothesis for short-term memory (1956), although in the original theory, chunking depended on preexisting units in memory such as letter groups that form words or acronyms. In CSTM, existing knowledge is

used to build structures that include new elements. Few sentences that we read or hear are recognized as a whole since most have never been encountered before, and yet a normal RSVP sentence is easily structured and retained long enough to be recalled immediately afterward. (I say "normal," because it would be easy enough to write a sentence with so many new elements and relations that it could not be successfully processed and retained in a single pass.) Similarly, with extensive practice a subject studied by Chase and Ericsson (1982) was able to develop coding schemes for structuring random sequences of digits, eventually expanding his digit span to over 80 digits. Clearly, the new structures were stored in long-term memory (except for the most recent digits that had not as yet been structured). Note that in this study the digits were presented at 1/sec; it is doubtful that the skill was sufficiently developed to have permitted structuring at a rate such as 10/sec.

If the structuring that occurs in conjunction with CSTM is in the same form or forms as the structuring characteristic of information in long-term memory, should one say that the resulting structure is "in" LTM? In terms of block diagrams of information flow, my answer is yes, but with the caveat that information in LTM can vary markedly in durability. Freshly structured information in CSTM must undergo a process of consolidation if it is to endure long enough to be recalled, and consolidation itself is a continuous variable. The details of our daily experience enter LTM for a time, but forgetting begins immediately and only the main incidents of the day are likely to be recallable the next day, and even less a week later. Information structured in CSTM is the leading edge of this negatively accelerated forgetting curve. In this view, conceptually structured experience is represented in a single memory system with a single consolidation process and a single forgetting function. (Adjunct memory systems such as phonologically based STM or imagery representations, with their own dynamics, may subserve the cuing or construction of conceptual memory.) However, the hypothesis that there is a single conceptual memory system may be oversimplified, given evidence from amnesics such as H.M. who seem to have intact CSTM (Potter, unpublished data) and yet totally forget new information within minutes (Milner, Corkin, & Teuber 1968).

How Does CSTM Interact with STM?

If conventional STM is largely irrelevant to most cognitive processing, which is carried out in conjunction with CSTM, then what is the role of STM and how does it interact with CSTM? By STM I mean conventional short-term memory as embodied in studies ranging from Miller's classic paper (1956) to Baddeley's (1986) working memory, in particular the articulatory loop system. (The visuospatial sketchpad and the central

executive, the other components of Baddeley's model, I do not discuss here: whether visualization is rapid enough to play a role in CSTM is doubtful, and the central executive is a residual memory device whose characteristics have only begun to be specified; see Baddeley 1996.)

This question about the relation between CSTM and STM is closely related to the assumption in the preceding section that more effortful and slower reasoning and thinking are carried out by essentially the same processes as in CSTM. The processes leading from one inference to the next may be the same, but more deliberate reasoning may require chains of inferences. The reasoner may need to pass over the material again and again before the needed conjunctions of ideas are made, and STM may be used to maintain relevant information in a retrievable form, reentering it into CSTM as such successive passes are made. The actual processes leading to a solution may be carried out stepwise in CSTM, with a longer latency between steps or with a longer chain of steps. On the other hand, it seems clear that one can use verbal STM or the visuospatial sketchpad to manipulate representations in the interest of solving a problem. (However, a surprising contrary result was recently reported by Butterworth, Cipolotti, and Warrington [1996], who studied an individual with a markedly impaired digit span who appeared to have normal ability to perform mental arithmetic.) There is evidence that patients with reduced memory spans may have trouble processing sentences with complex structures or temporary "garden paths," suggesting that sentence-processing does rely to some extent on the articulatory loop system—contrary to what Potter and Lombardi (1990) hypothesized (see above). But it is striking that many such patients adequately understand sentences, even though they paraphrase rather than report them verbatim (see Saffran and Martin's chapter in the present volume).

The performance of experts represents almost the opposite case: the prolonged training and practice required to become an expert in chess or medical diagnosis, or the playing of a musical instrument (Ericsson & Lehmann 1996), results in rapid recognition and action that is like "normal" CSTM, coupled with excellent memory. Ericsson and Kintsch (1995) have proposed that expert performance provides evidence for a long-term working memory (LT-WM) that keeps track of the status of a task in a particular domain, such as a chess game, permitting experts to play several games at the same time. In our terms, experts have developed the ability to structure and consolidate information in their special domain much more rapidly than nonexperts, as a result of their long training and current level of practice.

It is possible (as Ericsson & Kintsch 1995 suggest) that ordinary people (nonexperts) are in fact experts in the cognitive processing

of everyday life: perception and action in the (normal) environment, inferences about causal relations, language comprehension and production, reading, and so on. Independent of the question of innate endowment for performing these feats, it is evident that extensive practice in childhood is necessary for optimal performance in all these domains. Long practice can lead to a shift from slow, STM-bound processing to rapid CSTM processing (see the shift from a declarative to a procedural mode of processing, Anderson 1983).

Why Are the Most Convincing Demonstrations of CSTM All Visual (and Mostly Sequential)?

To examine the workings of the hypothesized CSTM, it is necessary to minimize the availability of other forms of memory, particularly conventional STM. Since the articulatory-phonological system that supports STM is derived from speech, auditory input virtually guarantees representation in this system. The phonological store provides an effective temporal buffer for short sequences of speech or other auditory input, so processing can be spread out over time. The buffer has some limitations, however. Using compressed speech (speech that is sped up without raising its pitch), Yntema, Wozencraft, and Klem (1964) showed that listeners became overloaded when more than 3 or 4 compressed digits were presented at a rate of 10/sec, forgetting many of them—just as viewers in our experiments with RSVP lists of words forgot most of them at that rate. But, unlike a visual presentation at 10/sec, compressed speech at that rate is markedly degraded; although it is possible to train a subject to recognize a finite set of one-syllable words such as digit names, it is difficult or impossible to comprehend sentences that use an unconstrained vocabulary at such high rates.

Sequential presentation is used to control the rate of processing and to examine continuous processing rather than processing of a single tachistoscopic stimulus. Because the viewer is obliged to attend to a succession of stimuli, performance reflects the capacity to process items and integrate them at a given rate. Normal reading rate is limited by the rate of eye movements and by ingrained reading habits, but RSVP reading demonstrates that people can process single sentences when reading more than twice as fast as they normally would. It is then possible (as in the paragraphs study, for example) to discover what processes begin to fail as reading is speeded up. In the present context, the advantage of using RSVP is that one is able to reveal phenomena that cannot be explained in the standard framework that includes only sensory memories, STM (including the visuospatial sketchpad), and LTM.

But does rapid visual presentation of sentences or other stimuli actually prevent material from entering standard STM? That is not yet

clear. On the one hand, there is strong evidence that RSVP readers do retrieve some phonological representation of the words they are reading (Bavelier & Potter 1992; Petrick 1981; Petrick & Potter 1979; and see Coltheart's chapter in this volume). On the other hand, concurrent articulation during RSVP reading does not appear to interfere with processing and immediate memory for the sentence (Potter 1984b), nor does it interfere with phonologically based repetition blindness between homophones such as ate/eight in RSVP sentences (Bavelier & Potter 1992). Besner and Davelaar (1982) have proposed that there are two phonological codes, the first of which is generated immediately, perhaps directly from the orthography, and contributes to lexical access. The second code is postaccess and constitutes the rehearsable phonological component of the articulatory loop model of STM. Only the second code is interfered with by articulatory suppression. If this hypothesis is correct, then it is likely that the early phonological code is the one activated during RSVP reading, whereas the late code, associated with STM, is not—but see Coltheart's chapter in the present volume for a different view.

Is CSTM Conscious?

At this point the question probably cannot be answered, because we have no clear independent criterion for consciousness other than availability for report. And, by hypothesis, report requires some form of consolidation, and therefore only what persists in a structured form will be reportable. Thus, while the evidence we have reviewed demonstrates that there is conceptual processing of material that was subsequently forgotten, it does not tell us whether we were briefly conscious of that material, or whether the activation and selection occurred unconsciously.

It seems unlikely, however, that multiple competing concepts (such as the multiple meanings of a word) that become active simultaneously could all be conscious in the ordinary sense, although preliminary structures or interpretations that are quickly discarded might be conscious. People do sometimes become aware of having momentarily considered an interpretation of a spoken word that turns out to be false, for example. And in viewing rapidly presented pictures, people have a sense of recognizing all the pictures but forgetting them. But such experiences seem to be the exception rather than the rule. Thus, I adopt the working hypothesis that much of CSTM activation and selection and structuring happens before one becomes aware: it is the structured result, typically, of which one is aware, which is why perception and cognition seem so effortless and accurate.

Conclusions

Evidence for conceptual short-term memory has appeared in a wide variety of tasks. The purpose of this chapter has been to review and discuss some of this evidence for early accessing of cognitive information and rapid selective structuring of that information. Rapid structuring can occur only if the material permits it and if the skills for discovering latent structure are highly practiced: for example, object and word recognition, lexical retrieval, sentence parsing, causal inference, search for a target, and the like. These are just the cognitive skills, each highly complex, that make comprehension seem trivially easy most of the time.

CSTM is the working memory that supports these processes, lasting just long enough to allow multiple options to be entertained before one is selected and the unused fragments evaporate, in most cases without entering awareness. The labored thoughts and decisions we are aware of pondering are a tiny fraction of those we make effortlessly. Even these worked-over thoughts may advance stepwise, by recirculating the data through CSTM until the next step occurs to us. We are aware of slowly shaping an idea or solving a problem, but not of precisely how each step occurs. More work will be needed to gain a full understanding of just what takes place in this largely preconscious stage of cognitive processing.

Notes

1. In most AB studies "lag 1" is defined as the item immediately following T1, whereas in many RB studies the immediately following item is considered to be at lag 0, and the next item is at lag 1. In this chapter I use "lag 1" to refer to the item immediately following T1 or C1.
2. The pictures were presented sequentially, and the viewer did not have to move his or her eyes to scan the picture; in one check on eye movements, we found that the viewers' eyes rarely moved in this task when the pictures were presented for 250 msec or less (Potter & Levy 1969).

References

- Allport, D. A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umiltà & M. Moscovitch (eds.), *Attention and performance XV* (pp. 421-452). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Altarriba, J., & Soltano, E. G. (1996). Repetition blindness and bilingual memory: Token individuation for translation equivalents. *Memory & Cognition*, 24, 700-711.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Arnell, K., & Jolicoeur, P. (1995). Allocating attention across stimulus modalities: Evidence from the attentional blink phenomenon. Poster session presented at the 36th annual meeting of the Psychonomic Society, Los Angeles, November.
- Arnell, K., & Jolicoeur, P. (1997). The attentional blink across stimulus modalities: Evidence for central processing limitations. Manuscript submitted for publication.
- Baddeley, A. (1986). *Working memory*. Oxford: Clarendon Press.
- Baddeley, A. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology*, 49A, 5-28.
- Bavelier, D. (1994). Repetition blindness between visually different items: The case of pictures and words. *Cognition*, 51, 199-236.
- Bavelier, D., & Potter, M. C. (1992). Visual and phonological codes in repetition blindness. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 134-147.
- Besner, D., & Davelaar, E. (1982). Basic processes in reading: Two phonological codes. *Canadian Journal of Psychology*, 36, 701-711.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18, 355-387.
- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 11, 717-726.
- Butterworth, B., Cipolotti, L., & Warrington, E. K. (1996). Short-term memory impairment and arithmetical ability. *Quarterly Journal of Experimental Psychology*, 49A, 251-262.
- Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. In G. H. Bower (ed.), *The psychology of learning and motivation* (vol. 16, pp. 1-58). New York: Academic Press.
- Chun, M. M. (1997a). Types and tokens in visual processing: A double dissociation between the attentional blink and repetition blindness. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 738-755.
- Chun, M. M. (1997b). Temporal binding errors are redistributed by the attentional blink. *Perception & Psychophysics*, 59, 1191-1199.
- Chun, M. M., Bromberg, H. S., & Potter, M. C. (1994). Conceptual similarity between targets and distractors in the attentional blink. Poster session presented at the 35th annual meeting of the Psychonomic Society, St. Louis, November.
- Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 109-127.
- Coltheart, M. (1983). Iconic memory. *Philosophical Transactions of the Royal Society of London*, B302, 283-294.
- Dooling, D. J., & Lachman, R. (1971). Effects of comprehension on retention of prose. *Journal of Experimental Psychology*, 88, 216-222.
- Downing, P., Kanwisher, N. C., & Potter, M. C. (1998). A crossmodal repetition deficit for RSVP and compressed-speech sequences. Manuscript under revision.
- Ericsson K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211-245.
- Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, 47, 273-305.
- Forster, K. I. (1970). Visual perception of rapidly presented word sequences of varying complexity. *Perception & Psychophysics*, 8, 215-221.
- 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, 115-125.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122-149.

- Kahneman, D., & Treisman, A. M. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davies (eds.), *Varieties of attention* (pp. 29–61). New York: Academic Press.
- Kanwisher, N. G. (1987). Repetition blindness: Type recognition without token individuation. *Cognition*, 27, 117–143.
- Kanwisher, N. G., & Potter, M. C. (1990). Repetition blindness: Levels of processing. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 30–47.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review*, 95, 163–183.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203–205.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104, 211–240.
- Lawrence, D. H. (1971a). Temporal numerosity estimates for word lists. *Perception & Psychophysics*, 10, 75–78.
- Lawrence, D. H. (1971b). Two studies of visual search for word targets with controlled rates of presentation. *Perception & Psychophysics*, 10, 85–89.
- Loftus, G. R. (1983). Eye fixations on text and scenes. In K. Rayner (ed.), *Eye movements in reading: Perceptual and language processes* (pp. 359–376). New York: Academic Press.
- 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.
- Lombardi, L., & Potter, M. C. (1992). The regeneration of syntax in short-term memory. *Journal of Memory and Language*, 31, 713–733.
- ✓ Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, 382, 616–618.
- MacKay, D. G., Abrams, L., Pedroza, M. J., & Miller, M. D. (1996). Cross-language facilitation, semantic blindness, and the relation between language and memory: A reply to Altarriba and Soltano. *Memory & Cognition*, 24, 712–718.
- MacKay, D. G., & Miller, M. D. (1994). Semantic blindness: Repeated concepts are difficult to encode and recall under time pressure. *Psychological Science*, 5, 52–55.
- ✓ Maki, W. S., Couture, T., Frigen, K., & Lien, D. (1997). Sources of the attentional blink during rapid serial visual presentation: Perceptual interference and retrieval competition. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 1393–1411.
- Maki, W. S., Frigen, K., & Paulson, K. (1997). Associative priming by targets and distractors during rapid serial visual presentation: Does word meaning survive the attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, 23, 1014–1034.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, 88, 375–407.
- McConkie, G. W., & Currie, C. B. (1996). Visual stability across saccades while viewing complex pictures. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 563–581.
- McKoon, G., & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 1155–1172.
- McKoon, G., & Ratcliff, R. (1998). Memory-based language processing: Psycholinguistic research in the 1990s. *Annual Review of Psychology*, 49, 25–42.
- Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1423–1442.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.
- Miller, M. D., & MacKay, D. G. (1994). Repetition deafness: Repeated words in computer-compressed speech are difficult to encode and recall. *Psychological Science*, 5, 47–51.
- Miller, M. D., & MacKay, D. G. (1996). Relations between language and memory: The case of repetition deafness. *Psychological Science*, 7, 347–351.
- Milner, B., Corkin, S., & Teuber, H. L. (1968). Further analysis of the hippocampal amnesic syndrome: 14-year follow up study of H. M. *Neuropsychologia*, 6, 215–234.
- Miyake, A., Just, M. A., & Carpenter, P. A. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neutral contexts. *Journal of Memory and Language*, 33, 175–202.
- Monsell, S. (1996). Control of mental processes. In V. Bruce (ed.), *Unsolved mysteries of the mind: Tutorial essays in cognition* (pp. 93–148). Hove, E. Sussex: Erlbaum (UK) Taylor & Francis.
- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Norris, D. (1986). Word recognition: Context effects without priming. *Cognition*, 22, 93–136.
- O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, 46, 461–488.
- Petrick, S. (1981). Acoustic and semantic encoding during rapid reading. Ph.D. dissertation, Massachusetts Institute of Technology.
- Petrick, S., & Potter, M. C. (1979). RSVP sentences and word lists: Representation of meaning and sound. Paper presented at the 20th annual meeting of the Psychonomic Society, Phoenix, November.
- Potter, M. C. (1975). Meaning in visual search. *Science*, 187, 965–966.
- Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 509–522.
- Potter, M. C. (1982). Very short-term memory: In one eye and out the other. Paper presented at the 23rd annual meeting of the Psychonomic Society, Minneapolis, November.
- 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. 413–437). New York: Academic Press.
- Potter, M. C. (1984a). Rapid serial visual presentation (RSVP): A method for studying language processing. In D. Kieras & M. A. Just (eds.), *New methods in reading comprehension research* (pp. 91–118). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Potter, M. C. (1984b). Articulatory suppression and very-short-term memory for sentences. Paper presented at the 25th annual meeting of the Psychonomic Society, San Antonio, TX, November.
- Potter, M. C. (1993). Very short-term conceptual memory. *Memory & Cognition*, 21, 156–161.
- Potter, M. C., Chun, M. M., Banks, B. S., & Muckenhoupt, M. (1998). Two attentional deficits in serial target search: The visual attentional blink and an amodal task-switch deficit. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 979–992.
- Potter, M. C., Chun, M. M., & Muckenhoupt, M. (1995). Auditory attention does not blink. Paper presented at the 36th annual meeting of the Psychonomic Society, Los Angeles, November.
- Potter, M. C., Kroll, J. F., & Harris, C. (1980). Comprehension and memory in rapid sequential reading. In R. Nickerson (ed.), *Attention and performance VIII* (pp. 395–418). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Potter, M. C., Kroll, J. F., Yachzel, B., Carpenter, E., & Sherman, J. (1986). Pictures in sentences: Understanding without words. *Journal of Experimental Psychology: General*, 115, 281-294.
- Potter, M. C., and Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, 81, 10-15.
- Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, 29, 633-654.
- Potter, M. C., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, 38, 265-282.
- 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., Stiefbold, D. R., & Moryadas, A. (1998). Word selection in reading sentences: Preceding versus following contexts. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24, 68-100.
- Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, 18, 849-860.
- Rayner, K. (ed.). (1983). *Eye movements in reading: Perceptual and language processes*. New York: Academic Press.
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8, 368-373.
- Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124, 207-231.
- Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, 2, 437-442.
- Sachs, J. S. (1974). Memory in reading and listening to discourse. *Memory & Cognition*, 2, 95-100.
- Sereno, S. C., & Rayner, K. (1992). Fast priming during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 173-184.
- ✓ Shapiro, K., Driver, J., Ward, R., & Sorenson, R. E. (1997). Priming from the attentional blink: A failure to extract visual tokens but not visual types. *Psychological Science*, 8, 95-100.
- Shulman, H. G. (1971). Similarity effects in short-term memory. *Psychological Bulletin*, 75, 399-415.
- Sperling, G., Budiansky, J., Spivak, J. G., & Johnson, M. C. (1971). Extremely rapid visual search: The maximum rate of scanning letters for the presence of a numeral. *Science*, 174, 307-311.
- Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645-659.
- Von Eckardt, B., & Potter, M. C. (1985). Clauses and the semantic representation of words. *Memory & Cognition*, 13, 371-376.
- Yntema, D. B., Wozencraft, F. T., & Klem, L. (1964). Immediate serial recall of digits presented at very high rates. Paper presented at the 5th annual meeting of the Psychonomic Society, November.