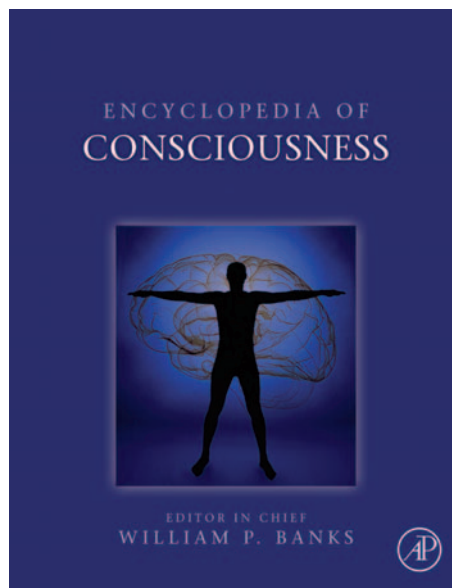


**Provided for non-commercial research and educational use.  
Not for reproduction, distribution or commercial use.**

This article was originally published in the *Encyclopedia of Consciousness* published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who you know, and providing a copy to your institution's administrator.



All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

<http://www.elsevier.com/locate/permissionusematerial>

Cowan N (2009), Sensory and Immediate Memory. In: William P. Banks, (Editor), *Encyclopedia of Consciousness*. volume 2, pp. 327-339. Oxford: Elsevier.

## Sensory and Immediate Memory

N Cowan, University of Missouri, Columbia, MO, USA

© 2009 Elsevier Inc. All rights reserved.

### Glossary

**Auditory stream segregation** – The tendency for two types of sound that are presented in an intermixed fashion to be perceived as coming from separate sources, or perceptual streams; this is more likely to occur at faster presentation rates.

**Continuous moment hypothesis** – The hypothesis that events are perceived together when they fall into the same sliding window of time; this window is continually updated.

**Discrete moment hypothesis** – The hypothesis that events are perceived together in nonoverlapping bins of time; any events falling into the same bin are perceived as simultaneous.

**Immediate memory** – Temporary information about a stimulus set, such as a word list or object array, that can be retained long enough to allow correct performance on a test shortly afterward, with no intervening interference or distraction. As used here, it includes both sensory memory and temporary memory for concepts.

**Perceptual memory** – A term that can be considered a synonym for sensory memory. See the introduction for possible differences.

**Perceptual moment** – Also sometimes called the psychological moment; the small amount of time in which two successive events seem simultaneous, which happens only when they are within about 200 ms of one another.

**Psychological present** – A period of up to a second or so, during which a single event seems to take place.

**Sensory memory** – Temporarily remembering how certain things look, sound, feel, taste, or smell.

**Short-term memory** – A term that can be considered a synonym for immediate memory. See the introduction for possible differences.

**Working memory** – A term that can be considered a synonym for immediate memory.

### Introduction to Concepts and Terms

Sensory memory is temporarily remembering how certain things look, sound, feel, taste, or smell. Immediate memory refers to information about a stimulus set, such as a word list or object array, that can be retained temporarily to allow correct performance on a test shortly afterward, with no intervening interference or distraction. Researchers differ in their definitions but, as used here, immediate memory includes both sensory memory and also the short-lived memory of concepts. The concepts can come either from stimuli or from thoughts.

What is distinct about immediate memory is that it is short-lived; it has a certain richness that fades with time or is eliminated by subsequent stimuli that cause distraction or interference. What is left when it is gone is a more impoverished, yet still critically important, long-term memory representation of sensations, events, and concepts from the past.

Researchers also vary in the particular terms they use. For some, a synonym for sensory memory is perceptual memory. Other researchers would think of perceptual memory to include identification of the objects that are being perceived (e.g., “I am seeing a chair” or “I am hearing a bird”), but would exclude that information from sensory memory. For some, synonyms for immediate memory are short-term memory or working

memory. Other researchers would think of short-term memory as referring only to the information automatically held, whereas working memory or immediate memory would include that information but also information held with the benefit of attention and mnemonic strategies, such as covert verbal rehearsal. We will circumvent these disagreements and stick to the terms sensory and immediate memory.

At any moment, you are likely to be aware of some of the information contributing to sensory and immediate memory, but not other information. Imagine that you are at a large gathering of friends in a large room, with refreshments. The vast field of chatter in the room impinges on your auditory system, but you cannot make out all of the conversations and, mostly, your awareness of this chatter recedes into the background. You are holding a conversation with one person, on financing a new automobile, and several other people nearby are talking about a topic that interests you, a new political sex scandal. It is impossible for you to attend fully to both of these things at once. As you start picking up more about the sex scandal, your understanding of what you are being told about automobile financing starts to wane, as you struggle to stay polite. You notice your conversational partner making a distressed facial expression as he sees your increasing boredom with the topic and, to allay his distress, you say something to demonstrate that you are in fact paying attention. You do not remember much of the content of the recent discussion, but you still can recall the last few words that he said, which you repeat in an interested, questioning tone of voice. The general chatter in the room, the snippets from the sex conversation, and the monologue about car finance are all part of your sensory, perceptual, and immediate memory.

### **Attention and Immediate Memory**

A very important distinction is between information to which you pay attention and information that you ignore. The ignored information does not disappear immediately. Think of the last few words of the car financing conversation that you could repeat although you were not fully listening.

Similarly, while you are engrossed in a novel, someone may ask you what time it is. You ask them to repeat what they said but, before they do repeat it, you are able to recall what was asked. How did you do that? You pulled it from sensory memory into a more conceptual immediate memory.

Donald Broadbent studied this kind of issue in the 1940s and 1950s, helping to establish a new field called cognitive psychology, the study of perception, memory, and thought processes. He and his colleagues in England were interested in some practical issues, such as how a pilot could land a plane safely when the radio instructions arriving through headphones are mixed with instructions being delivered simultaneously to other pilots landing other planes. That problem was eventually solved through advances in radio technology. In the psychology laboratory, however, the problem was studied using a technique called dichotic listening. One message was played to one ear and a different message in a different voice was played to the other ear concurrently. The task was to shadow, or repeat, everything presented to one ear. From time to time the tape was turned off and the subject was asked to repeat any information presented in the other ear, the ignored message. Subjects were stumped. They often could remember a few words from the end of the ignored message, but that was all. This led to a distinction between attended and unattended aspects of immediate memory. So, for example, you take in stimuli from all sensory modalities and some of the sensations outlast their stimuli. You can practically hear the last notes of a symphony for some seconds after they have been played. Following a flash of lightning on a dark night, your memory of the sight of twisting trees lingers on. Usually, though, there is such a wealth of stimulation that you cannot attend to all of it. A few sensations are selected for further processing, memorization, and understanding. They might be the actual words and ideas from a conversation, or a single canvas in an art museum.

Broadbent's theory was essentially that stimuli that are not relevant to one's current concerns are filtered out of attention and awareness. It is relatively easy to filter out undesired stimuli on the basis of physical cues, as when one ignores a certain person's voice or picks only red berries, not green ones. It is more effortful, but still possible,

to filter out nondesired stimuli on the basis of their meanings, as in a room of chatter when one ignores all conversations that are not related to particular subject, or when one goes into the garage to pick up several tools that are needed for a particular carpentry job while leaving others behind. In the latter type of selection, any kind of mental preoccupation can easily result in a failure of selection. You may then tend to hear a loud but irrelevant conversation, or you may pick up the wrong tools.

At one point early in the history of cognitive psychology, it was proposed that people actually take all stimuli into the mind for further processing, and are limited only in the ability to respond to so much at once. That late-filter theory was based partly on a further dichotic listening study by Neville Moray in 1959, in which the subject's own name was presented in the ignored message. It was found that some people notice their name, suggesting that the message was not actually filtered out. This finding was often highlighted in textbooks for many years, albeit without further support. (Instead, the field seemed to focus on visual stimuli.) Much later studies in the 1990s did confirm that this registration of one's own name sometimes occurs.

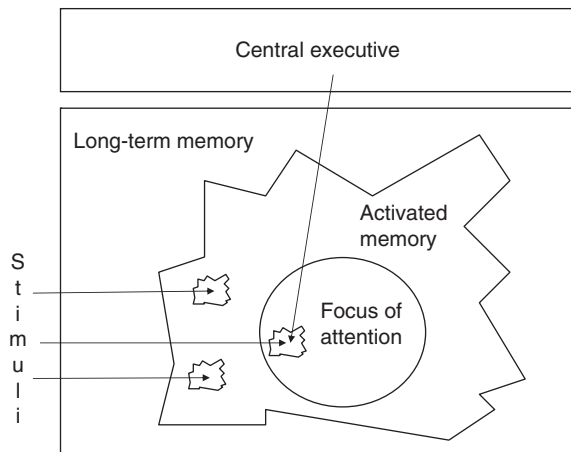
A study by Andrew Conway and his colleagues in 2001 discussed Moray's result and suggested that this type of finding, noticing one's name in an ignored message, does not actually support the late-filter theory. Another explanation is that Broadbent was essentially correct in suggesting an early attentional filter, but that some people fail to maintain a strong task set so that the information actually attended does not reliably match the information that was supposed to be attended. Conway and colleagues tested subjects on a complex immediate-memory test called operation span, in which arithmetic problems were to be solved along with memorization of a word presented after each problem. All of the words were to be recalled after the last arithmetic problem was solved. The operation span test, like certain other complex working memory span tests, correlates with intelligence rather well, and it correlates with the ability to pay attention. For example, recent research by Michael Kane shows that individuals with low spans experience more involuntary mind-wandering when they are trying to pay

attention than do high-span individuals. If dichotic listening really does involve selective filtering and subjects' attention sometimes wanders, it could wander off of the message that was supposed to be attended and onto the message that was supposed to be ignored, or it could be split between the two messages. That should happen more frequently among low-span individuals. Consistent with this interpretation, Conway and colleagues found that 65% of the subjects in the lowest quartile of operation span noticed their names, whereas only 20% of the subjects in the highest quartile noticed. So Moray's result does not strongly challenge Broadbent's filter theory after all. The results for noticing one's name are consistent with an early filter theory.

Johnston and Heinz, in 1978, used a selective listening task to show that people can pay attention either to sensory features, as in the early filter theory, or to conceptual features, as in the late-filter theory. However, it takes much more effort to pay attention according to the late-filter theory. In this experiment, there was a secondary task, reaction time when a light was presented. Reaction time to the light was about the same when the task was carried out alone or along with selective listening based on a sensory feature, voice quality. During selective listening to one of two messages about different topics spoken in the same voice, however, reaction time to the light was much slower. This indicated that it takes effort to pay attention to conceptual features, whereas paying attention to physical features is easy and natural as in the early filter theory.

## A Schematic Model of the Information Processing System

Using the information that has been presented so far, you can begin to form a concept of how the brain's information processing system acts. One way to conceive of the system is shown in [Figure 1](#), from Nelson Cowan's writings. Incoming stimuli all make contact with elements in long-term memory that become activated, and these activated elements of memory include sensory memory. However, not all activated elements enter the focus of one's attention, just as the entire room full of



**Figure 1** A concept of the immediate-memory system, which includes both sensory and conceptual features. Activated memory and the focus of attention are embedded processes that both contribute to immediate memory. For further support see Cowan's 1995 and 2005 books.

chatter mentioned above, although it can be heard, cannot all be comprehended. The Central Executive represents mental processes that you use to impose your own will in terms of what information to focus your attention on and what to do with that attended information, as when you decide to concentrate on understanding one speaker while ignoring other conversations. However, other factors help control the focus of your attention, such as abrupt noises, sudden changes in the lighting, and attractive displays on television. These can either help your central executive, as when a speaker uses dramatic changes in voice intonation to get across a point, or can hinder your central executive, as when you try to listen to a monotonous speaker during a thunder storm. The information that is attended receives further processing in any case, resulting in a collection of concepts that accompany the sensory images in one's conscious mind and mold one's interpretation of these images. (Both sensations and concepts can count as activated elements from long-term memory.) The tendency for the central executive to be able to control which sensory stimulus channels to attend and which to ignore is roughly equivalent to Broadbent's filter. In the illustration, paying attention on a sensory or perceptual basis (as in picking only the red berries) is like staying on one stimulus input arrow, whereas paying attention on a

conceptual basis (as in finding which tools one needs) is like hopping from one arrow to another.

Not everything in immediate memory can be an activated set of representations from long-term memory, of course. New information is represented by a new set of links between already-existing features of some sort. For example, if you saw a blue apple, it would probably be the first link of the blue feature and the shape associated with apples. You might also receive a new name for it. These new links between a shape, a color, and a combination of phonemes would coexist in the focus of attention and would produce a new long-term memory representation.

One reason to use a schematic diagram like this one is that the brain is complicated. Any one of these components included in **Figure 1** could correspond to a twisted tangle of tree-like neural structures distributed widely throughout the brain. The schematic model is at a higher level of analysis. By analogy, one can delineate the executive, legislative, and judicial branches of the US government when, in fact, the three branches of government are represented by complex and overlapping webs of people, buildings, and regulations. Nevertheless, there are known neural correlates of these elements of mind. The central executive processes rely a great deal on frontal lobe structures (behind the forehead). The parietal areas, more toward the upper middle of the head, are very much involved in the focus of attention; damage to those areas can result in patients who are unaware of one part of space (unilateral neglect) or are unaware that they have a serious paralysis or other disability (anosognosia). Recent brain research shows that parts of the parietal lobes are especially active when a challenging amount of information is being saved in immediate-memory tasks. The temporal lobe areas (behind the sides of the head) include the hippocampus and other neighboring areas that are especially important in retaining information and forming new long-term memories, and diverse areas represent different kinds of memories that become electrically active when the ideas that are represented are activated by stimulation or thought processes.

The characteristics of sensory and immediate memory that make them particularly interesting stem partly from their limits. The discussion of

what these kinds of memory can do is intricately related to what they cannot do. In the remainder of this article, two types of limits will be discussed. The first is time limits in activated memory and the second is capacity limits in the focus of attention.

### Time Limits of Activated Memory

It is possible to demonstrate directly that there are limits in the activation of information from long-term memory, and this temporarily activated memory forms the basis of the human information processing scheme shown in [Figure 1](#). These are time-related limits, although the exact limit seems to depend on the nature of the stimulation. Some concepts related to human awareness may be based on this temporarily activated memory. These will be discussed first, and then more direct tests of the time limits of activation will be discussed.

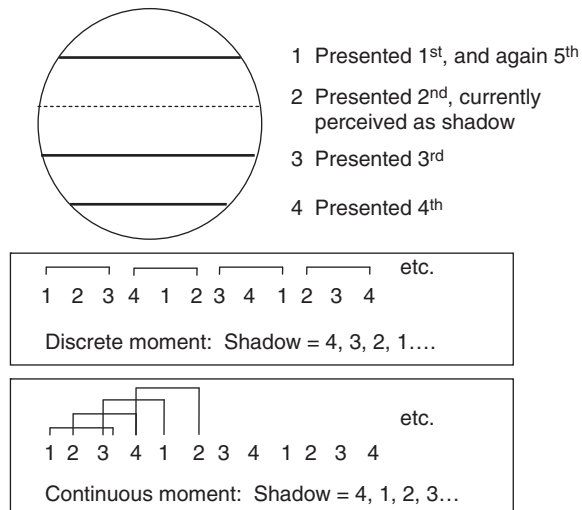
There are two concepts related to time limits and awareness that are of special interest, called the perceptual moment (also sometimes called the psychological moment) and the psychological present. The perceptual moment is the small amount of time in which two successive events seem simultaneous, which happens only when they are within about 200 ms of one another. As an example, if you move your finger back and forth as rapidly as you can, it will visually appear as if your finger is in two places at once (as well as being smeared in between these two places). Why does this occur? One way of explaining it is that the neural processes representing each visual viewpoint outlast the actual stimulus and are still ongoing when the next visual viewpoint arrives. This, in fact, is how movies are perceived. Another way of explaining it is to say that there is a visual sensory memory or afterimage of the stimulus that seems like a continuation of that stimulus after it is gone, and that the two afterimages are combined in our awareness into a single stimulus. If there were more time between the stimuli, one afterimage would fade before the next one was formed and the two viewpoints would not appear simultaneous. A great deal of research suggests that this neural explanation and this psychological explanation are fundamentally compatible and that both are apt. The

perceptual moment depends on the type of stimuli that are presented but typically ends within a couple hundred milliseconds.

In contrast to the fleeting perceptual moment, the psychological present is a longer period of time of a second or so, during which a single event seems to take place. The rapid beats of a drum roll may be grouped together to form a single event, whereas slower beats of a drum may be perceived as separate events. Whereas, for the perceptual moment to be the same for two stimuli, ongoing perceptual processing must overlap, the psychological present has a looser criterion; the second stimulus need only be presented while the first is still vividly recalled. It is as if there are two phases of temporarily activated sensory features in memory: a vivid afterimage that seems as if the stimulus is containing, lasting for several hundred milliseconds, and then a vivid recollection of the sensory events, for a second or so. Like the latter, William James in the late 1800s used to think of primary memory (in this chapter, immediate memory) as the trailing edge of the conscious present.

Many studies could be used to demonstrate these concepts. We will focus on a couple of them that seem especially thought-provoking. In 1968, D.A. Allport carried out a study to distinguish between two varieties of the perceptual moment hypothesis. In the discrete moment hypothesis, the stimuli are accumulated into one moment until some fixed amount of time; then subsequent stimuli are accumulated into another moment until a similar fixed amount of time; and so on, just as each sheet of paper in an office could be stamped with a date and multiple sheets that arrived at different times would share the same date stamp. In the continuous moment hypothesis, though, events would be viewed as if through a sliding window (i.e., sliding in time). First the window might integrate events 1, 2, and 3; then the oldest event 1 would fade from the window, to be replaced by a new event 4; and so on.

To distinguish between these hypotheses, Allport developed a simple but ingenious test using an oscilloscopic display, as shown in [Figure 2](#). (Actually, the display included 12 lines but the point can be made more simply using four lines as shown here.) The lines would be presented one



**Figure 2** Allport's experiment to distinguish between discrete and continuous varieties of the perceptual moment. The top part of the figure illustrates a particular moment of perception in the situation in which Lines 1–4 were presented one at a time, in order, repeatedly at a fast rate. Line 2 happens to be perceived as a shadow at the moment depicted here, but the perceived shadow location moves as the display progresses. The proposed mechanism is illustrated according to the discrete moment hypothesis in the middle of the figure (with brackets showing which lines four consecutive moments would include), and according to the continuous moment hypothesis at the bottom of the figure (with brackets showing which lines four overlapping, temporally sliding moments would include).

at a time in the repeating order 1, 2, 3, 4, 1, 2, 3, 4, . . . , and so on with no break. The subjects were to adjust the rate of line presentation until what was perceived was all but one of the lines at the same time. (If the rate were any faster, all of the lines would be perceived at once or, if the rate were any slower, fewer lines would be perceived at once.) With the rate so adjusted, the line that was not perceived at any one moment was called a shadow and the change over time in which line was seen as a shadow was called shadow movement. Then different predictions about shadow movement could be made for the two types of perceptual moment.

For the discrete moment, as shown in the middle of the figure, the assumption was that the succession of lines was parceled out into nonoverlapping moments of a certain duration. That process would result in shadow movement in the opposite direction from the actual movement of lines. If the first

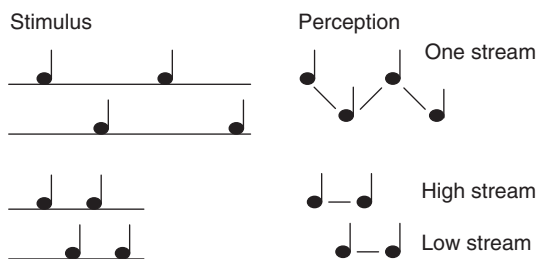
discrete moment were long enough to include Lines 1, 2, and 3, then 4 would be the shadow. The next discrete perceptual moment would open up just in time to capture Line 4, and also the next presentation of Lines 1 and 2, but then it would close before it could include the next presentation of Line 3, which would go in the next window instead. Thus, the shadow has moved from Line 4 to Line 3 and continues in that manner, in a direction opposite from that of the line presentations. In contrast, in the continuous moment hypothesis, shown in the bottom of the figure, the shadow movement would go in the same direction in which the lines were presented because the temporally sliding window always includes the most recently presented three lines and leaves out the least recently presented line. The evidence from the experiment unequivocally supported the continuous moment hypothesis, in which our events are grouped together according to a sliding window of perception. This sliding window can be viewed as sensory memory, in which the least recent line always was the first to fade from view.

Albert Bregman has carried out a line of research on a phenomenon that can be seen as demonstrating the power of the psychological present, called auditory stream segregation. Suppose you hear a simple series of two tones in alternation: high–low–high–low in pitch, and so on. How will the series sound? That depends on various factors, but one of these is the time between tones, as shown in [Figure 3](#). When the series is presented slowly, the mind links together adjacent tones into a single, coherent perceptual stream going up and down. When the series is presented more quickly, it is the tones of the same pitch that seem to be grouped together, into a high-pitched stream and a second, low-pitched stream. In fact, it is difficult to hear exactly when the high and low pitches were presented relative to one another; there is no clear perceptual organization across streams. Tones in each stream share the grouping principle of similarity. When the series is considerably slowed down, however, what may be different is that successive tones of the same pitch are no longer within the same psychological present, and so they cannot be grouped together. In terms of memory, it may be that the vividness of one tone has faded by the time

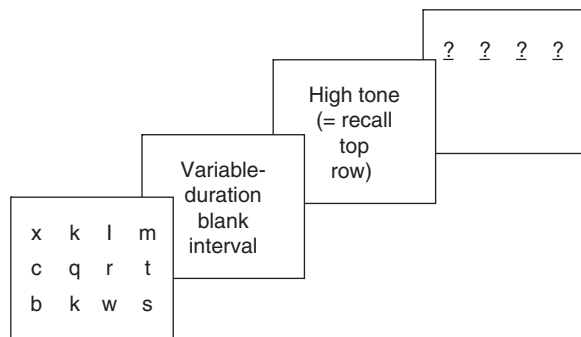
a similar tone is presented, so that they no longer seem to belong together in time.

Time-dependent activated memory does not rest primarily on notions of the perceptual moment and the psychological present, but has repeatedly been examined more directly through a wide variety of experimental procedures. One of these, by George Sperling, was published in 1960. He carried out a number of experiments. A typical one is illustrated in Figure 4.

The first panel of the figure shows an array of 12 letters that was presented very briefly, with different letters on every trial. In one kind of trial (not the kind shown in the figure), the task was simply to recall all 12 letters by writing them down. This was called the whole report procedure, and it showed that practiced adults could recall only about 4 of the 12 letters in their correct locations. Sperling realized, though, that he did not know the basis of this limit. It could be that more items than this were held in an activated form of memory but that they disappeared from memory before they could all be written down. What Sperling did to address this problem was to develop the



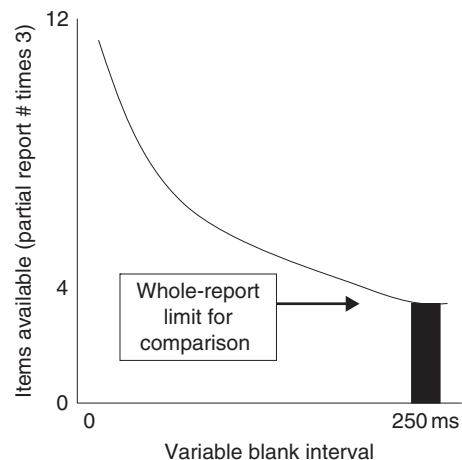
**Figure 3** The phenomenon of auditory stream segregation, as in Bregman's work.



**Figure 4** Sperling's partial report procedure.

partial report procedure, shown in Figure 4. Subjects learned that they should respond in a manner dictated by a tone presented after the array. If the tone was high, the task was to recall the top row; medium, the middle row; or low, the bottom row. Because the tone was not presented until after the array, when the array was presented the subject had to retain all three rows. However, after the tone was presented, the task became simply to recall one row, which did not exceed the number of items that subjects could recall in the whole report situation. The result would indicate how much information about the cued row was held in memory at the time that the cue arrived. Given that the other rows had to be held at the same time, the total amount of information available in memory when the tone cue arrived could be calculated by multiplying by 3 the average number of items recalled in partial report.

The outcome is shown in Figure 5. It turned out that the result depended on the amount of time between the array and the tone cue. If the tone was presented very soon after the array, almost all items in the array were still available for recall. However, if the tone cue was delayed for a quarter second (250 ms), it was of almost no value. The number of items available for recall was then no greater than the whole-report limit of about four items. This suggested that all of the items in the array were held in memory at first, but that they faded from memory within about 250 ms. Other experiments showed that the cue had to indicate a physical feature, such as the row of the array to



**Figure 5** Results of Sperling's whole and partial report procedures.

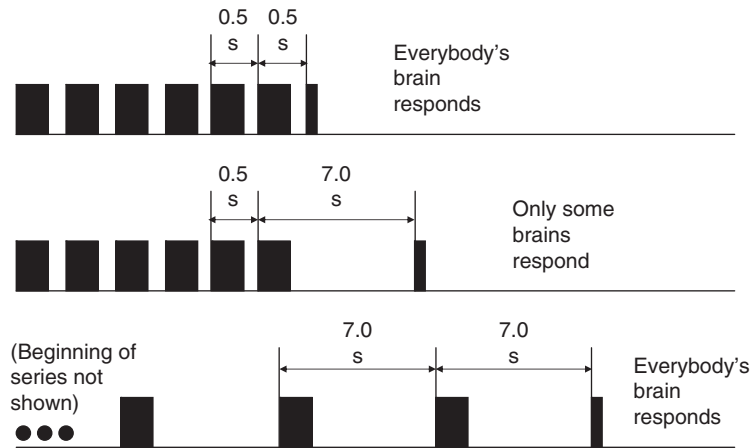
recall, and not a more abstract, conceptual feature, such as whether to recall letters or digits from an array that included both types of symbols. This suggests an explanation that is entirely consistent with Donald Broadbent's explanation of the dichotic listening results. A large amount of sensory information is held for a brief time, but only a small subset of that sensory information can be shifted to the focus of attention before it is lost. It can be selected for attention and further processing on the basis of physical cues (such as voice quality or direction, or written letter color or location) more easily than it can be selected on the basis of conceptual cues (such as the topic of speech, or the category of the written symbols).

One possible discrepancy between auditory and visual sensory memory in these studies is that the duration of auditory sensory memory appeared to be several seconds, whereas the visual sensory memory only appeared to last a fraction of a second. In a review that Cowan published in 1988 and discussed further in a 1995 book, an alternative interpretation of the literature on sensory memory was suggested. It appeared that, all information considered, there are two separate phases of sensory memory: a vivid mental afterimage experienced while perceptual processing of the items continued for about 250 ms, and a longer-lasting, vivid sensory recollection for several seconds. There has not been a great deal of discussion of this issue (i.e., whether there are sensory-modality-specific memory durations or, instead, two phases of sensory memory in all modalities), but it remains one on which researchers do not seem to agree. The two phases of sensory memory discussed in the 1995 book, if they do exist, would appear to underlie the perceptual moment and the psychological present, respectively, as we have discussed.

These experiments leave open the question of whether nonsensory, conceptual information that is activated also has a limited time period of persistence. This is not an easy issue to examine, because sensory information ordinarily may have to be drawn into the focus of attention before conceptual features can be perceived. Then it becomes possible to rehearse the information by repeating it covertly (mentally and silently) or simply by continuing to attend to it. One can prevent this rehearsal by presenting distracting

tasks, but these can cause interference as well. In 1973, Michael Watkins and his colleagues presented lists of words followed by tones that either could be ignored or, in a different type of trial, had to be identified by button press, and found that memory was lost over time – especially during the first 3 s of a 20-s period – only when the tone information had to be identified. On the other hand, in 2004, Stephan Lewandowsky and his colleagues took a different approach and found that memory for letters was not lost over time when the words had to be recalled either slow or fast, regardless of whether subjects had to recite a word over and over to prevent covert verbal rehearsal. Some details of both studies, and of other studies in this literature, seem to be such that the issue of the loss of conceptual activated memory features over time remains unresolved.

Even if a loss of activated memory over time is observed, it has to be asked whether it is the amount of time itself that is important, or the way in which that time is perceived relative to recent events. This is clear in a study of brain function based on electrical signals at the scalp, or event-related potentials, by István Winker and his colleagues in 2001. They presented series of tones that remained identical for six tones, with the seventh tone sometimes shorter than the others. Even though the subjects are busy reading during this kind of procedure, so that the tones are ignored, subjects' brains still respond to tone changes if the unusual tone (the deviant) is presented shortly after the other tones (the standards). This mismatch negativity response is said to occur because the brain saves a sensory memory of the standard tones and compares each new tone to that standard representation. **Figure 6** shows three different stimulus arrangements and indicates the results. When all tones in the series were a half second apart, all subjects' brains responded to the change in tone duration. However, when the standard tones were one half second apart and the deviant did not occur until 7 s after the last standard, some subjects' brains still responded, whereas others' brains did not. For the subjects whose brains did not respond, one might think that a memory of the standards was lost by 7 s. However, another interpretation is that their brains no longer considered those standard tones to serve as a fair, current



**Figure 6** The procedure that Winkler and colleagues used to examine why tone memory declines over time.

comparison. After all, the standards in each series formed a tight group in time and the deviant was not part of that group. Verifying that interpretation, all subjects' brains responded when all tones were 7 s apart, as in the third row of the figure.

To summarize this discussion of temporarily activated memory, it can include sensory or conceptual features from long-term memory that have been activated by recent events or thoughts. The conceptual features are much more likely to be activated only after information is drawn into the focus of attention, but then the focus of attention might shift elsewhere and leave behind activated conceptual features that are unattended. For sensory and conceptual activated features, it is clear that there is a time limit to activation but, after all these years of research, it is not yet clear what kind of time limit that is. All of the currently available information might be compatible with the notion that the critical time interval may be derived by the brain relative to the pace of events. At least, that may be the case for the longer form of activated memory that we have linked to the psychological present. It is usually within the range of a few seconds but it is hard to pin down exactly.

Within that possibly flexible duration of the psychological present, there appears to be a shorter duration of about 250 ms that may be determined by the time it takes for perceptual processing of a single stimulus to be completed, which therefore may be a more precisely fixed duration. It corresponds to a vivid sensory afterimage in the brain and to the perceptual moment.

### Capacity Limits of the Focus of Attention

One of the functions of the focus of attention is to overcome time limits. Suppose you want to check the oven in 3 min. Without an external timer, you have no choice but to invest a considerable amount of effort attending to the oven so that your mind does not wander on to other things. The fundamental reason why this is effortful is that there are competing stimuli and competing thoughts, and not all of them can be attended at once. That is, there is a capacity limit to the focus of attention.

One dramatic result of the capacity limit is what Dan Simons and others have called inattentive blindness. You can demonstrate it by playing a trick on a friend in the office. Wait until they turn around for a moment and then remove a small item from plain sight on their desk, such as a stapler. If you do this discretely chances are that, when your friend returns attention to your direction a moment later, he or she will not notice the missing object. Provided that the item was not recently attended, despite being well within the perceptual field of view, its disappearance will not attract notice. Simons and his colleagues demonstrated inattentive blindness dramatically by having one experimenter stop a pedestrian to ask directions. In a staged event, two other confederates of the experimenter holding a door in a horizontal position walked between the experimenter and the hapless pedestrian who was asked directions, obscuring their view of each other. Secretly, on the side of the

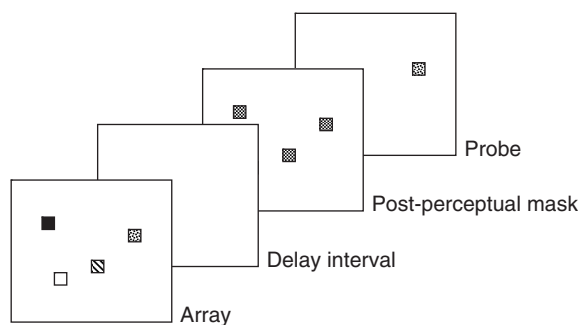
door across from the pedestrian, the experimenter grabbed the door and was replaced by one of the confederates who had been holding the door. When the two people with the door moved along out of the way, the confederate acted as if he was the original experimenter. In many cases, the pedestrian did not notice that there had been a change in the person to whom he was giving directions before the interruption, and continued to give directions. He had been attending to his directions, and not primarily to the exact features of the person receiving those instructions.

Sperling's partial report procedure shown in [Figure 4](#) could be considered an inattentive blindness procedure because attention is insufficient to encode into the focus of attention all of the items from the array held in visual sensory memory. In 1997, Steve Luck and Ed Vogel simplified that procedure so that it would not require that multiple items be recalled, in order to get the most sensitive measure of what items are in working memory. These might be thought of as the items encoded into the focus of attention. A version of that procedure used recently by Jeffrey Rouder and his colleagues does an excellent job of estimating items conceptually held in working memory. The procedure is illustrated in [Figure 7](#). An array of small, colored squares is briefly presented. (Each pattern in the figure represents a different color.) An interval of about 1 s then elapses, giving the subject plenty of time to encode the array into attention. Then a masking display is presented with a set of multicolored squares in the same location as the colored squares in the previous array. This serves to eliminate any remaining sensory memory. What remains are any conceptual

aspects of working memory that the subject has retained even though the sensory memory features have been overwritten by the mask. The masking display also serves to remind the subject just where the squares were on the screen. This is followed by a single colored square, and the task is to indicate whether this square is the same color as the square in that location within the original array. If it is not, it is a color that was not found in the array. The probe remains until an answer is given.

This procedure produced results that conformed very well to a simple mathematical model to calculate capacity. The model included an assumption that on a certain percentage of trials, the subject was not paying attention at some point. That happened on 12% of the trials according to the model. Each subject was said to have a certain capacity  $k$  and, when he or she paid attention,  $k$  items were retained in working memory. When the array item in the same location as the probe was in working memory, the subject was assumed to know whether there was a color change or not. If the item was not in working memory, either because of inattention or because capacity was exceeded, the subject was assumed to guess. In different trial blocks, different proportions of trials involved color changes to manipulate the guessing rate. The estimate of  $k$  stayed fixed across guessing rates and across different numbers of items in the array, as it should according to the mathematical model. The mean  $k$  was 3.35 items, consistent with previous estimates of working memory capacity such as those reviewed in Cowan's 2005 book.

Scott Saults and Cowan recently showed that this kind of working memory examined by Rouder and his colleagues is not limited to visual information. They set up four loudspeakers surrounding the subject and played four different spoken digits at the same time, each in a different voice, while also presenting an array of four or eight colored squares on the computer screen. The subject was responsible sometimes for the visual information only, sometimes for the auditory information only, and sometimes for both modalities at once. Provided that a postperceptual mask was presented to eliminate sensory memory in both modalities, the visual and auditory information traded off. In particular, subjects remembered about 3.5 visual items when they did not have to retain auditory items also, and



**Figure 7** A version of the visual working memory procedure by Steven Luck and Edward Vogel.

they remembered about 3.5 items total when they had to remember the auditory along with the visual. The verbal spoken items displaced some of the non-verbal visual items in working memory, which seems to be evidence that the information being stored in a capacity-limited manner is conceptual and abstract rather than specific to a sensory modality.

The memory in these procedures we have been discussing seems less than what was discussed in a famous article by George Miller, who, in 1956, suggested that there is a “magical number seven plus or minus two,” describing how many items people can recall. This estimate is roughly true, but it does not seem to reflect the most basic type of immediate memory. Instead, it seems to reflect a process whereby the items in working memory are grouped together to make larger items. (That grouping process was in fact one of Miller’s main points.) When one reads the telephone number 356-4129, one tries to form a group of three items followed by a group of four. When the groups are formed, each group may take up only one slot in working memory. It is these smaller groups that may reflect the basic capacity limit, with the larger limit occurring as several smaller groups are entered into working memory together. Anders Ericsson and his colleagues have made the point that, with enough practice, at least some individuals manage to raise their memory span from seven items to the astonishing level of eighty or more items, but only for the type of material being practiced (such as digits). They do it by learning to form large chunks based on knowledge they already have, such as memorized athletic records. More precisely, chunks of up to about four items are formed and then these chunks are combined to form even larger, higher-level chunks.

The capacity of working memory increases in childhood and then decreases again in old age, as Cowan, Moshe Naveh-Benjamin, and their colleagues have demonstrated recently. Also, as Randall Engle and his colleagues have emphasized, tests of working memory capacity do a good job of discriminating between young adults with more versus less aptitude on complex cognitive tasks and intelligence tests. Yet, we do not understand what it is that distinguishes someone with a high span from someone with a lower span. One theory is that individuals with a high span simply have more slots in working

memory than do lower-span individuals. However, an alternative theory is that individuals with a high span are able to pay attention better, and therefore are more likely to fill their working memory capacity with relevant items as opposed to thoughts that are irrelevant to the task that has been assigned.

Jim Gold and his colleagues found evidence for the slots theory in distinguishing schizophrenic individuals from normal controls. In one experiment, for example, they presented arrays with colored items of two shapes and usually tested subjects on one of these shapes. Occasionally, they would test subjects on the other shape. In this way they could tell how well individuals were filtering out the usually irrelevant shape (by examining the advantage in performance on the usually tested shape over the other shape) and they could tell how much individuals encoded into working memory (by examining the total of capacity  $k$  for the two shapes added together). Surprisingly, the schizophrenic individuals filtered just fine but were below normal on total capacity. In contrast, Vogel and his colleagues have done similar testing using event-related brain potentials and have argued that high- and low-span normal individuals differ primarily in the ability to filter out the irrelevant information. We know of areas in the brain that specialize in holding working memory information (certain areas in the parietal lobes) and areas in the brain that specialize in filtering out irrelevant information (the basal ganglia). It stands to reason, then, that the available slots in working memory and the ability to fill these slots appropriately and efficiently are two important, related, but nonidentical aspects of individual and group differences.

## Conclusion

Immediate memory, or its close cousin working memory, is critical for human thought and awareness of the environment. It is, as William James said, the trailing edge of the conscious present. The conceptualization of this memory as shown in [Figure 1](#) has proved to be useful in understanding the mind. Activated features of memory, within which sensory memory usually predominates, are collected by our brains, apparently for all stimuli, without the need to pay attention. Even though the

information that is activated remains so only briefly, it provides a rich backdrop from which information can be drawn selectively into the focus of attention for further processing in a timely manner. Conceptual features that come out of this selective processing can be added back into the pool of activated memory. There are some key questions that remain unanswered regarding the principles that produce the limits in activated memory and the focus of attention, but the scientific endeavor is briskly marching toward some tentative answers.

### Acknowledgments

This work was supported by NIH Grant R01 HD-21338.

See also: Visual Experience and Immediate Memory.

### Suggested Readings

- Allport DA (1968) Phenomenal simultaneity and the perceptual moment hypothesis. *British Journal of Psychology* 59: 395–406.
- Bregman AS (1990) *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.
- Broadbent DE (1958) *Perception and Communication*. New York: Pergamon Press.
- Cowan N (2005) *Working Memory Capacity*. Hove, East Sussex, UK: Psychology Press.
- Engle RW, Tuholski SW, Laughlin JE, and Conway ARA (1999) Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General* 128: 309–331.
- Ericsson KA, Delaney PF, Weaver G, and Mahadevan R (2004) Uncovering the structure of a memorist's superior "basic" memory capacity. *Cognitive Psychology* 49: 191–237.
- Gold JM, Fuller RL, Robinson BM, McMahon RP, Braun EL, and Luck SJ (2006) Intact attentional control of working memory encoding in schizophrenia. *Journal of Abnormal Psychology* 115: 658–673.
- Johnston WA and Heinz SP (1978) Flexibility and capacity demands of attention. *Journal of Experimental Psychology: General* 107: 420–435.
- Kane MJ, Brown LH, McVay JC, Silvia PJ, Myin-Germeys I, and Kwapił TR (2007) For whom the mind wanders, and when: An experience-sampling study of working memory and executive control in daily life. *Psychological Science* 18: 614–621.
- Lewandowsky S, Duncan M, and Brown GDA (2004) Time does not cause forgetting in short-term serial recall. *Psychonomic Bulletin & Review* 11: 771–790.
- Luck SJ and Vogel EK (1997) The capacity of visual working memory for features and conjunctions. *Nature* 390: 279–281.
- McCollough AW and Vogel EK (2008) Your inner spam filter. *Scientific American Mind*, June/July: 32–35.
- Miller GA (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63: 81–97.
- Simons DJ and Levin DT (1998) Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review* 5: 644–649.
- Winkler I, Schröger E, and Cowan N (2001) The role of large-scale memory organization in the mismatch negativity event-related brain potential. *Journal of Cognitive Neuroscience* 13: 59–71.

### Biographical Sketch



Nelson Cowan (PhD, 1980, University of Wisconsin) is a Curators' Professor at the University of Missouri. He has been interested in the brain and mental processes every since he was in high school, and this interest developed further in college at the University of Michigan. His research, funded by the National Institutes of Health since 1984, focuses on working memory, selective attention,

and the childhood development of these processes. He is the author of *Attention and Memory: An Integrated Framework* (1995, Oxford University Press) and *Working Memory Capacity* (2005, Psychology Press). He has served as an associate editor of three journals, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *Quarterly Journal of Experimental Psychology*, and *European Journal of Cognitive Psychology*. As of September 2008, he is on the governing board of the Psychonomic Society and is the president of the experimental psychology division of the American Psychological Association. Nelson is married and has three grown children; his wife, Jean Ispa, is a professor of human development and family studies.