Survey Paper on Visual memory

By
Nate Gertsch and Joel Ossher

CS 231

1 Introduction

Visual memory differs from other forms of visual perception. While other forms of visual perception can be traced to stimulus, the trigger for memory is internal. While other forms of visual perception can be examined directly, memory is nebulous and needs careful experiments in order for us to examine its properties. Most importantly, while other forms of visual perception allow us to make sense of the stimuli that bombard us every instance, memory allows us to create meaning from the chain of instances and be able to function in real life. It is not an exaggeration then that memory is of the highest importance.

Historically, research on human memory has been focused in the verbal domain. The reason for this is primarily methodological; it is much easier to present verbal materials to subjects, either is visual or auditory form. Much more is therefore known regarding how people remember letters, words and sentences than objects and scenes. This has also resulted in the theories of human memory being biased towards verbal information. However, some of the results have generalized well to visual memory, which will be discussed in much greater detail.

The formalized study of verbal memory began with the work of Hermann Ebbinghaus in the late 1800s. He proposed an undifferentiated human memory system, one where information was only "forgotten" due to interference of newer memories. By far the most popular viewpoint, there were still a few dissenters. William James argued instead for the existence of two kinds of memory: **primary memory**, which lasts for 10s of seconds, and **secondary memory** which can last for years. Primary memory would be used, for example, to dial in a phone number after looking it up. It could be maintained nearly indefinitely by explicit rehearsal, but would otherwise quickly fade away.

The debate over the exact form of human memory came to a head in the 1960s, when new experiments supported James' view of the existence of a differentiated memory system with duration of 30 seconds or less. This briefer memory system is now usually called **short-term memory (STM)**, while the longer system is known as **long-term memory (LTM)**. Over the last 40 years the research on human memory has converged on a three-tiered system of memory. Thousands of experiments utilizing many different procedure types have attempted to tease apart the contributions of each of these individual systems.

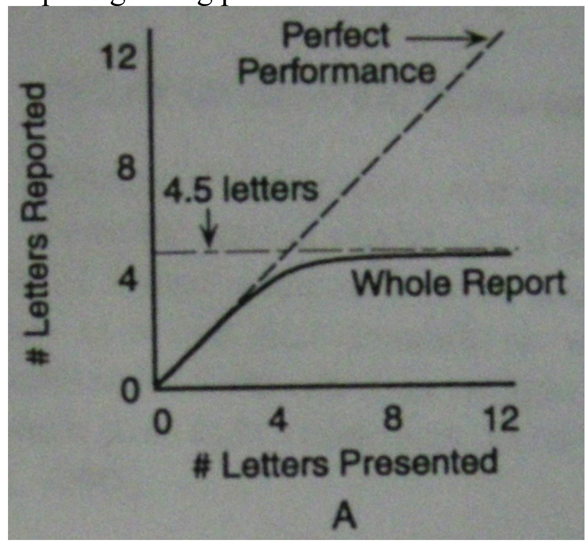
The shortest form of memory are the **sensory information stores (SIS)**, which are extremely brief "buffer" memory stores for information in very specific modalities, such as visual or auditory. Short-term memory is a slightly longer-lasting system with a distinctly limited capacity, used to store information that is currently being processed. Long-term memory does not appear to have an explicitly limited capacity, nor an explicitly limited duration, allowing us to remember a lot for a very long time.

When comparing these three different systems of memory, we will focus on 5 primary distinguishing attributes. *Duration*, the length of time that memories last within the system. This is the property after which memory systems are often named. *Content* covers the kind of information the memory system can store, both the distinction between verbal and auditory as well as differences within a specific modality. *Loss* relates to how the information is lost from memory. The two most common causes are time (loss from autonomous decay) and interference (loss from conflicting new memories). *Capacity* is how much information the memory system can store. Lastly, *maintenance* regards how information can be voluntarily refreshed, if at all. Some memory systems support explicit rehearsal, while others do not. Now we can look at the memory systems individually.

2. Iconic Memory

The discovery of iconic memory, a form of very brief high capacity memory, happened in the 1960s when George Sperling became interested in studying the nature of brief visual memories. Previous research had sought to determine the **span of apprehension**, the number of letters a person could perceive in a single, brief visual presentation. Subjects were presented with a grid of as many as 20 letters for less than 100ms, and asked to report as many of them as they could. As can be seen in the accompanying figure, when presented with only a few letters subjects' performance was quite good. But as more letters were presented their performance peaked, and they could typically only identify four or five items.

Sperling ran himself in the experiment, and found that he too could only report four or five letters. This didn't mesh with his visual experience, however. He felt that for a short period of time he could see all of the letters, but that before he could finish reporting them, this memory would fade away. He reasoned that the problem was not with his memory, but instead with the **whole report procedure** itself. When asked to report all the letters they could perceive, the memory had faded away before reporting could be completed. He needed a way to test how many letters could be identified without requiring a long procedure.



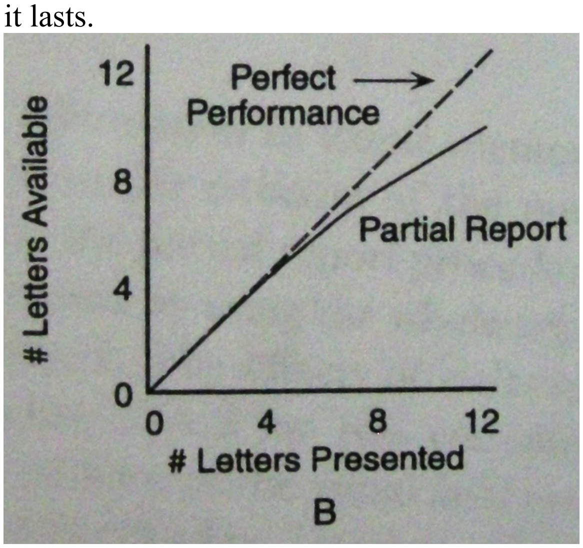
2.1 Partial Report Procedure

Sperling developed an ingenious new method, called the **partial report procedure**, to test his hypothesis. Similarly to the standard span of apprehension tests, Sperling (1960) presented subjects with a grid of letters, this time arranged into three rows. After the brief presentation of the visual stimuli, a tone was played to indicate which row of letters to report. A high tone indicated the top row, a medium tone the middle row, and a long tone the bottom row. Given the order of presentation, subjects could not use the tone cue to selectively attend to a specific row of letters. Therefore any letters they could successfully report must have been available to the memory system. By only asking for one third as many items to be reported, the time required to report was significantly reduced.

In order to estimate the number of letters stored in iconic memory, Sperling took the number of rows and multiplied it by the number of letters correctly identified per row. The reasoning was that if a subject could report all the letters in a target row, and that target row was only cued after presentation of the display, then all of the letters in all of the rows must have been available in memory.

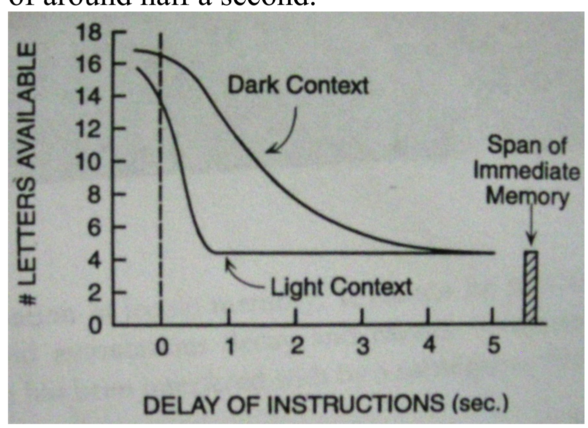
When Sperling ran the procedure, he found that he was indeed correct, and that iconic memory was much better than had initially been thought. The figure to the right shows that subjects only began running into difficulties when approaching the same four letter report limit as with the whole report procedure.

Averbach & Coriell (1961) provided more accurate estimates of the size of iconic memory by further refining the whole report procedure. Instead of cuing the correct row with a tone, they instead tried using a visual bar marker to indicate a specific letter to report. After presenting the grid of letters, a line was drawn under the location of where a letter had been, indicating that only that letter should be reported. By completely eliminating the performance penalty due to multiple report, they got average levels of performance as high as 16 of 18 letters presented. If one makes allowances for inattention and memory lapses, it appears that iconic memory may be able to hold virtually complete visual information while it lasts.



2.2 Duration of Iconic Memory

In order to test out the duration of iconic memory, Averbach and Sperling (1960) introduced a delay between the end of the letter display and the onset of the cuing tone. The rationale was that if the tone was delayed sufficiently then the information in iconic memory would have fully decayed, degrading performance from the levels seen in the partial report procedure to those of the whole report procedure (as we will see later, performance in the whole report procedure is thought to be governed by a different form of memory). The figure to the left shows the results of this. The dark versus light context indicates whether the pre-stimulus and post-stimulus fields are light and dark. As can be seen, in a light context the duration of iconic memory is less than a second, while in a dark context it can last for maybe a couple seconds. Given that in our typical experience pre- and post-exposure conditions are in a light context, our iconic memory has the duration of around half a second.



2.3 Content of Iconic Memory

While it is clear that iconic memory stores visual information, the exact form of that visual information is not immediately apparent. On approach to exploring the content of iconic memory is to study the effectiveness of different types of cues on partial report procedure performance. The idea is that a cue will only increase performance above the levels of the whole report procedure if iconic memory explicitly stores information of that type. Given that spatial cues (e.g. tone cuing a row, bar marker cuing a single letter) increase performance, we can conclude that iconic memory stores spatial position.

von Wright (1970, 1972) performed a series of experiments that illustrate the differential effectiveness of cue type. He arranged eight colored items into two rows of four. Four of the items were letters while four were digits. Four of them were red, while four were black. In each trial, one of two tones cued which set of four items to be reported. There was one condition for spatial location, another for category (letter versus digit), and another for color. All the cuing conditions were exactly the same, except for the type of information to which they directed attention.

The results of this experiment showed that both spatial information and color were effective cues. Both significantly increased performance for the partial report procedure. Category, however, was completely ineffective. Other experiments have shown that both size and shape are effective cues. From this we can conclude that location, color, size and shape are all features represented in iconic memory, while category is either not represented or cannot be used to access this memory.

2.4 Maintenance of Iconic Memory

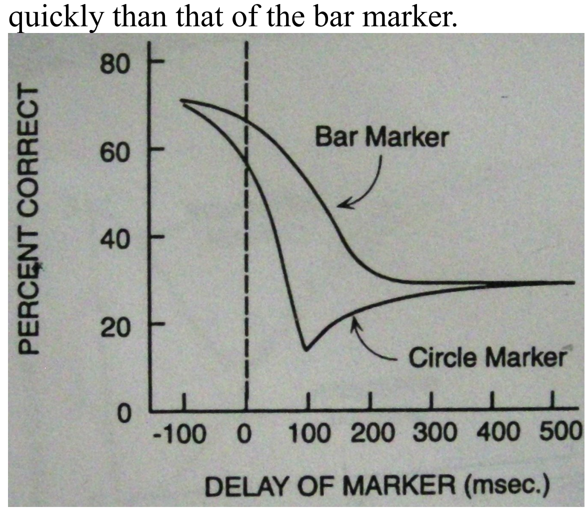
The information in iconic memory is either lost through autonomous decay or interference from subsequent items (the results from the duration experiments suggests that both methods play a role, as both time and light/dark context play a role). There does not appear to be a way to voluntarily maintain any information in iconic memory, short of continuing to look at the stimulus.

2.5 Loss of Iconic Memory / Masking

As was just mentioned, although autonomous decay seems to be the primary method in which information is lost from iconic memory, the duration experiments clearly indicated that interference also plays a role. The name given to this sort of interference is **masking**.

One form of masking was discovered completely by accident by Averbach and Coriell (1961) when they were experimenting with different kinds of spatial cues for partial reporting of single letters. They had the idea to cue a particular letter by surrounding it with a circle. Under certain delay conditions, this had the effect of completely erasing the perception of the letter inside. Subjects reported seeing an empty circle! This effect is known as **metacontrast masking** or **erasure**.

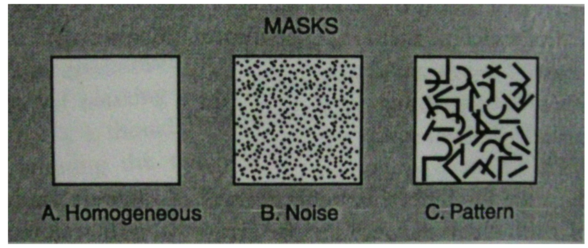
Further experimentation showed that metacontrast masking was most effective if the circle was presented 100 ms after the target display. Averbach and Coriell showed subjects two lines of letters, alternating between a circle and a bar market, and varying the time between the onset of the cue and the target display (from 100 ms before to 500 ms after). The results shown in the figure to the right indicate that while performance started off the same, the circle marker performance degraded much more quickly than that of the bar marker.



A further oddity regarding the phenomenon is that if the masking circle is itself masked by presenting an even larger circle, then the target letter is unmasked. This result implies that the larger circle effectively erases the smaller one, removing whatever obstacle existed for perception of the target letter. To properly understand this phenomenon, a more detailed exploration of masking is warranted.

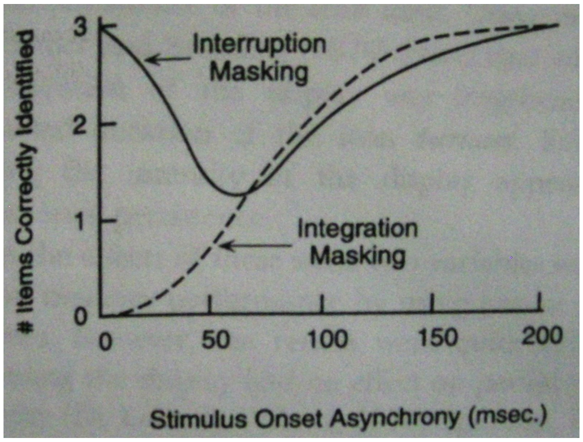
The timing at which the mask is presented relative to the target is important, giving us two different types of masking: **forward** and **backward masking**. As the names imply, forward masking is when the mask is presented before the target, and backward masking is when the mask is presented after the target. The exact form that the mask can take also varies. As seen in the figure to the left, **homogeneous** masks are uniform flashes of light, while **noise masks** are a random configuration of dots. Lastly, **pattern masks** are a randomly distributed collection of segments of more complex figures similar to the target itself.

Turvey (1973) has argued that there are at least two distinct mechanisms involved in masking that result in two different forms. **Integration masking** occurs when the target and mask images are, in effect, added together in memory, giving the effect of one single fused image. Such integration obviously occurs if the two images are optically superimposed (presented at the same time), but integration masking can also occur if the mask is present in very close temporal proximity to the target, not necessarily overlapping. Thus the effect from integration masking should be strongest when the target and the mask are presented very close to one another, the effect diminishing the greater the temporal distance.



The effects of integration masking can be seen in the dashed curve in the image to the right. These shows the results from an experiment where subjects had to report as many letters as they could from a three-letter row in the presence of a pattern mask displayed at twice the brightness of the target display. Both the target and the mask were presented for 10 ms. The x-axis shows the variation in **stimulus onset asynchrony (SOA)**, the time between the onset of the target and the onset of the mask. At 0 ms SOA, true optical integration of the target and the mask occur, making any letter identification completely impossible. These effects persist with an SOA of 16 ms, even though the target and the mask are no longer overlapping in their presentation. As the SOA increases, the effect of the integration masking continued to diminish until reporting is near perfect at 200 ms.

The solid curve shows a masking effect with a very different structure. At 0 ms there is no masking effect, but as the SOA increases the effect worsens, until at 50 ms it peaks and begins to diminish again. This curve was obtained by presenting the same stimulus as in the integration masking scenario, but instead lowering the brightness of the mask to one half the brightness level of the target instead of double. Because of the absence of a masking effect at 0 ms, we can conclude that the integration masking effect is minimal. This suggests a different form of masking is at work, which is called **interruption masking**. The metacontrast masking discussed earlier appears to be a form of interruption masking.



Integration masking is thought to occur very early on in visual processing, while interruption masking is thought to occur later. An important piece of evidence in support of this theory comes from a series of studies involving interocular masking. Homogeneous masks are generally thought to cause integration masking, as their effect in interruption masking scenarios are minimal. When presented to the same eye as the target, the integration masking effects of a homogeneous mask are devastating. However, when applied to a different eye as the target, they have no effect. This indicates that the integration masking effect likely occurs before are V1 of the visual cortex, as this is where the images from the two eyes are integrating. The lack of an interocular integration masking effect is in marked contrast to what occurs with pattern masks presented at interruption masking timing. In these cases, even presenting the mask to the non-target eye still results in the masking effect. This indicates that the interruption masking occurs more centrally (at or after the level of V1) than integration masking, and supports the existence of two different mechanisms.

2.6 Persistence of Iconic Memory

So far we have discussed in detail the duration in memory of a briefly flashed visual display. But what is the effect of this memory on conscious perception. Is it possible to report items from iconic memory because we can still see them past when they have been presented? This may seem like the obvious reason, but there is cause to doubt it. This evidence comes from studies involving the effect of stimulus duration and brightness on visual persistence and cued memory recall.

The visual persistence of a flashed display can be measured by a rather clever experimental procedure. Subjects were asked to adjust the timing of a tone so it would coincide with the perceived onset of offset of the displayed items. The difference between these two times thus corresponds to the perceived length of the display. Subtracting the actual duration of the stimulus from this perceived length yields the length of visual persistence. When using this technique, Haber and Standing (1970) found that as the duration of the display was lengthened, the perception of duration decreased. Similarly, increasing the brightness of the display also decreased perception of the duration. However, when the effects of these alternations were measured on cued visual recall using the partial report procedure, it was found that they had either no effect, or a positive one. Thus it is dangerous to equate phenomenal duration with the length of iconic memory.

Much is known about the functioning of iconic memory, but much less is known as to the reason for its existence. What purpose does it serve? One theory is that it is used to process visual information during a saccade. There is evidence that processing of the previous visual display continues during a saccadic eye movement, but it would be going a bit far to claim that iconic memory exists to facilitate this process.

Another possibility is that iconic memory serves to aid in motion processing. Some computational models of motion processing suggest comparison between two separate regions over a temporal delay, which iconic memory might help.

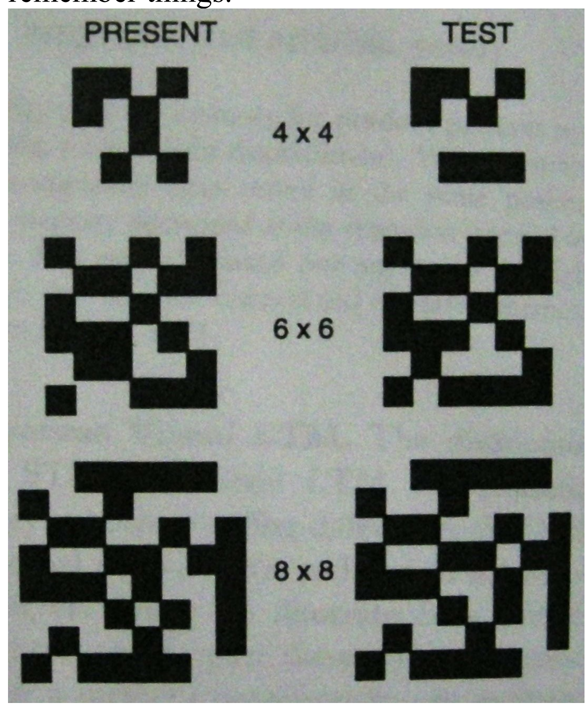
It's also quite possible that iconic memory is just a by-product of some other mechanism, and does not have any specific functional significance. One such theory suggests that iconic memory results from residual retinal activity in the rods. Rods are much more sensitive to light than cones, and so function under low-light conditions. Plus they have a slower neural decay rate than cones, and so will continue to fire even after a brief visual stimulus is presented. As tempting as this hypothesis sounds, further experimentation has revealed that partial report performance is intact even under conditions in which rods could not contribute. So to date the functional role of iconic memory is not known.

3. Short-Term Memory

As we have seen, iconic memory is a very short term form of visual memory, usually lasting less than one second. We clearly have the ability to store long-term visual memories, as evidenced by our ability to recognize people and places years after we have seen them. It would seem necessary, then, for there to be an additional memory system between iconic and long-term memory, enabling us to bridge the gap between our fleeting remembrances and our effectively permanent storage. Experimental evidence suggests the existence of such a form of memory, which is called **visual short-term memory**.

3.1 Differentiating Visual STM and Iconic Memory

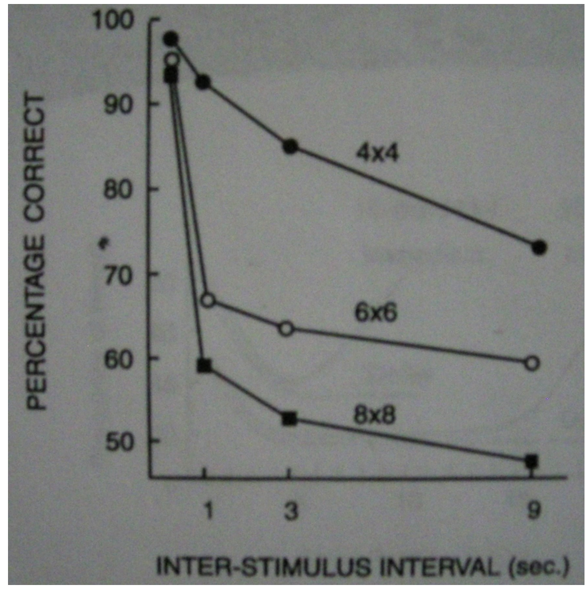
Given that iconic memory is thought to be precategorical (recall the failure of category as a cue in the partial report task), a seemingly obvious experiment to prove the existence of visual STM would be to show someone a meaningful display and then ask them to remember it more than one second later. This would work, except that when presented with categorizable displays, subjects categorize them. And once categorized, their minds have access to a variety of related information, which no longer limits them strictly to the visual domain, so it is impossible to know if they are using visual memory in order to remember things.



The main solution to this problem in the study of visual short-term memory has been to use meaningless visual materials. William Phillips went with this approach, filling matrices of varying sizes (from 4x4 to 8x8) with black and white squares. In his experiment (1974), he presented subjects with one matrix, and then, after a retention interval varying from 0 to 9 seconds, presented them with a second matrix that was either identical or differed by exactly one square. Subjects were asked to make a same/different judgment of the second (test) pattern with respect to the first (target) one.

As can be seen in the image to the right, at very short retention intervals performance was nearly perfect. This is what could be expected when the memory task is mediated by iconic memory. But at longer retention intervals, while performance does indeed decline dramatically, it remains well above chance levels, especially for the smaller matrices. Phillips suggested that this provided clear evidence for a form of visual memory that lasts well past the duration of iconic memory.

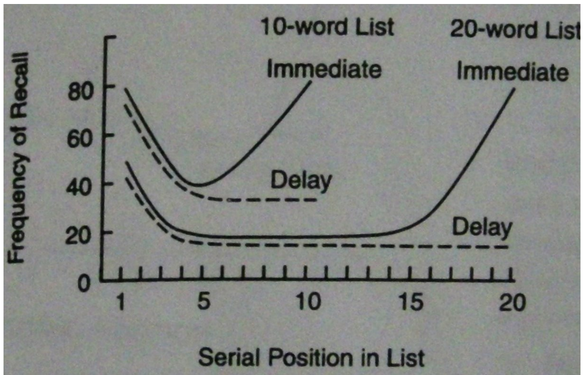
Phillips did further research exploring the properties of visual short-term memory. In one set of experiments, he investigated the effect of spatial position. Iconic memory is generally thought to be sensitive to spatial positioning in absolute retinal coordinates, so changing the location of the image should greatly disrupt performance that relies upon iconic memory. Phillips thought, however, that it was possible that visual short-term memory might be in object relative reference frames, in which case performance should be unaffected by movement. This is exactly what Phillips attempted (1974). After presenting the standard matrix target pattern, the test pattern was either presented in the same location or in a different one. Performance under these various conditions can be seen in the figure to the left. When matrix positions were not changed, performance at very short delays was again excellent, as to be expected, while quickly declining as the delay was increased. However, when matrix positions were changed, the retention curve was essentially flat, indicating that the initial contribution from iconic memory was completely eliminated while short-term memory was unaffected. So we can conclude that visual short-term memory represents information in terms of relative position.

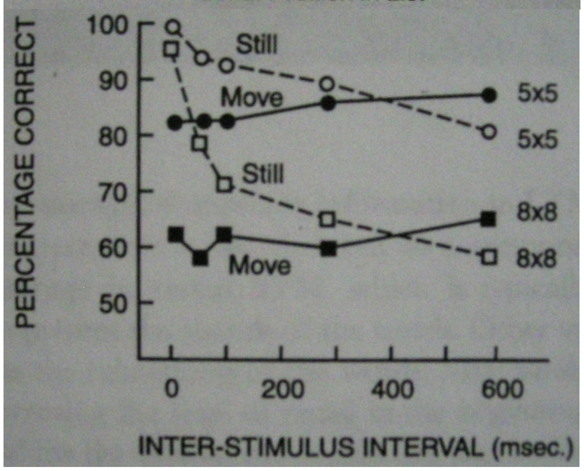


3.2 Differentiating Visual STM and Visual LTM

In order to properly understand how the difference between visual STM and visual LTM was justified, it is worth a brief digression into how verbal memory researches first differentiated these mechanisms in the verbal domain, using a memory paradigm called **free recall**.

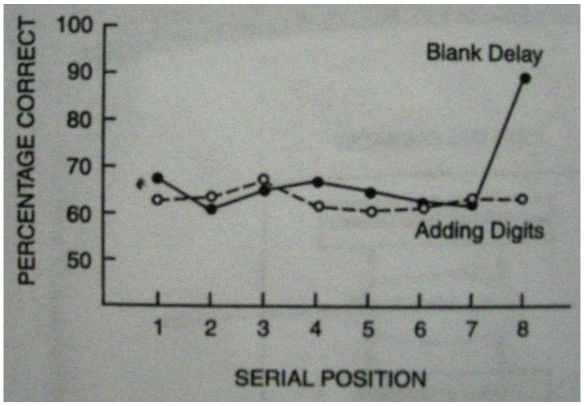
In a standard verbal free recall experiment, subjects are presented with a list of words (usually between 10 and 30 of them) in sequence, and then asked to recall as many of them as possible. The results are summarized a **serial position curve**, a plot of the probability of correctly recalling an item versus its position in the list. The accompanying figure shows a fairly standard serial position curve. They begin with a relatively high probability, which drops to a constant level in the middle region. The probability of correct recall then rises steeply at the end.





Memory theorists attribute the shape of this curve to two different mechanisms. The high levels of recall for the last few items in the list is deemed the **recency effect**, which is believed to reflect retrieval from short-term memory. Recall of the rest of the list is attributed to long-term memory, with the beginning being elevated due to the **primacy effect**. The reason that recall from different portions of this list is thought to arise from different memory systems is that they respond quite differently to a simple change of procedure. In the basic free recall experiment, subjects are asked to begin recalling the words immediately after their presentation is completed. Under a different condition, however, subjects are forced to wait 30 seconds before performing any recall. Furthermore, during the delay subjects are asked to perform a cognitively demanding distractor task, such as counting backwards by 3s from an arbitrary three digit number. With the addition of this distractor task, the recency effect disappears entirely, and the recall curve remains flat after the primacy effect. Other manipulations can affect the early parts of the list without changing the recall of the recency effect. For example, slowing the presentation time serves to improve recall in the early part of the list, presumably because subjects have a longer time in which to encode the word into their LTM.

Phillips and Christie (1977) investigated the application of this experimental technique to the visual domain, in order to illustrate the distinction between visual short-term and long-term memory. They decided to use the same black and white square matrix stimuli which had been used in the earlier forced choice recognition tasks. As with the verbal free recall experiments, subjects were presented with a series of 8 matrices in sequence. However, rather than being asked to do a recall task (which would be extremely difficult in the visual domain), subjects were instead given the same forced choice same/different judgment task as with the earlier experiments. As can be seen in the accompanying figure, a recency effect was indeed observed, but only for the last item presented. Furthermore, when given a cognitively demanding distractor task, the recency effect disappeared entirely, just as in the verbal experiments. They also tried manipulating the presentation rate, which, when increased, depressed performance in the early parts of the display while keeping the recency effect intact. From this we can conclude that visual STM is indeed different from visual LTM, in much the same way that verbal STM and LTM differ.



3.3 The Visuo-Spatial Scratch Pad

The visuo-spatial scratch pad is a theoretical memory construct that shares many properties with visual STM, proposed by Baddeley and Hitch (1974) as part of their model of working memory. Working memory is closely related to the standard concept of short-term memory, but differs slightly in its focus. Short-term memory is a memory storage system whose main purpose is to bridge the gap between the very short-term sensory information stores and long-term member. Working memory instead is meant to emphasize the role that short-term memory plays in virtually all cognitive processing tasks.

Working memory is proposed to have an important internal structure. It contains a central executive that controls general cognitive processing and problem solving. Linked to these central executive are a number of slave memories whose purpose is to provide a buffer for recently accessed information in that sensory modality. Thus visual information is stored in the visuo-spatial scratch pad, while auditory information is stored in the articulatory loop.

There is some experimental evidence that supports this multistore system for memory. Logie (1986) performed an experiment in which subjects were asked to use one of two memory strategies: either forming visual images or rote rehearsal. He then measured the effect of different forms of distractor stimuli. In the visual imagery condition, he found that visual distractors degraded performance while irrelevant words did not. The results were reversed in the other condition.

3.4 Transsaccadic Memory

When one's eyes perform a saccade, moving from one location to another, the world appears stable. Things do not appear to jump around, and the new view integrates well into the old. Rayner, McConkie and Erlich (1978) suggested that there might be an integrative visual buffer that superimposes the contents of successive fixations in order to achieve this effect. As reasonable as this might seem, experimental evidence has been to the contrary.

  Irwin, Yantis and Jonides (1983) reasoned that if such an integrative form of memory existed, then subjects should be able to superimpose the contents of multiple fixations in order to perceive the composite of two images. They tested this by presenting subjects with 12 dots from a 5x5 grid, forcing them to saccade, and then presenting another 12 dots from the same grid at the same spatial location (but now a differerent retinal location). Subjects were asked to identify the location of the missing dot. When the subjects are not forced to saccade between presentations (and hence the images are retinally superimposed), their performance is quite good. In the eye movement condition, however, subjects reported no effects of visual fusion and performance was extremely low.

Further experiments by Irwin took the partial report procedure discussed earlier and forced a saccade between the presentation of the stimuli and the cue. In this condition, the effect of iconic memory was entirely eliminated, and performance fell to levels consistent with visual short-term memory. This indicates that whatever transsaccadic memory we have is in reality just our visual short-term memory.

So why does the world look so stable when we move our eyes? Because the world generally is stable, and our minds know this and so give us that perception. As it turns out, if, during a saccade, the background is changed, subjects will not even notice. So this impression of stability persists even when it is not the case.

3.5 Conceptual Short-Term Memory

All of the work on visual short-term memory far discussed so far focuses on meaningless visual stimuli, so as to eliminate any bleed-over effect from other memory modalities. But another interesting area of study is how long it takes to perform categorization on visual stimuli, and what things can interrupt this categorization. The first studies on this appeared in the 1970s when an experimental paradigm known as **rapid series visual presentation (RSVP)** was developed.

In RSVP subjects are shown a series of pictures (from 10s to 1000s), each presented for a short period of time (from 100 ms to 2s). Subjects might be asked either to detect the presence of a specific type of image (e.g., two people having a picnic) or remember images for later old/new recognition tasks. At the slower presentation rates, subjects are able to both detect and recognize conceptually described pictures. At faster presentation rates, they are still able to detect the pictures, but are no longer able to recognize them later. It appears that people are able to comprehend pictures shown to them for 100 to 200 ms, but that that is not long enough to form a stable memory trace.

Further experimentation has shown that people's memories for pictures is quite good when they are shown for only 100 ms, even when followed by a noise mask. The recognition problem arises when multiple pictures are shown in succession. The appears to be a form of conceptual masking, in which the presentation of additional pictures within 500 ms of the previous one interrupts processing of the first, preventing it from being remembered.

4. Long-Term Memory

While short term memory has many interesting properties, it has a limited range. Our iconic memory fades in under a second while our short term visual memory only holds information for up to ten seconds. To visualize anything that we haven't seen in the last ten seconds, we require something that does not fade and is still available after some time has passed. We require long term memory. Before we begin with the details of visual memory, we will examine long term memory in general.

4.1 Types of long term memory

Memory theorists hold that long term memory is not a separate unique system, but part of a general system that includes other kinds of informational storage as well. This means that we will need to distinguishes between the different types of memory and see if these separate types have visual equivalents or not. The first type of memory that is generally held to exist is semantic memory. This is the place where concepts are stored and abstract information makes its home. For example: The concept that triangles have three sides is stored here. The concept that parrots are a kind of bird is another kind of semantic memory. Interestingly enough, there is a visual equivalent to this kind of memory. The simple shapes, called geons [1], are basic enough that some believe that they are always stored in memory and that we use these geons to make sense of our visual perceptions.

The second kind of memory that is generally held to exist is that of procedural memory. If you've ever wondered how a musician remembers all the motions to create the notes in a complicated piece of music, this is the answer. Procedural memory stores the HOW information of memory. This kind of memory lets us remember the steps necessary to perform an action and to repeat it. Many of the things that we take for granted are forms of procedural memory and are the kinds of information that is difficult to articulate. This kind of memory allows us to form skills and execute common tasks without having to rediscover every action. What does this have to do with visual memory? Ullman [2] believed that visual activities such as examining a picture, reading a book and so on where made up of these simple tasks called "visual routines". These small learned actions functioned the same way that basic physical actions did and formed a basis for understanding how we process visual information.

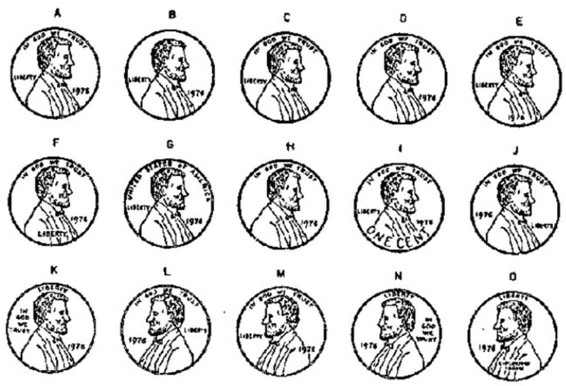
The third kind of memory is the most iconic (as in the normal sense of the word, not to be confused with iconic memory) kind of memory. This is the kind of memory that lets us tell boring stories about events that happened a long time ago. It allows us to remember our 5th birthday party and sequences of events that occurred outside the range of the short term memory. As we shall see, this kind of memory has some very interesting properties and can be prone to misinformation if not treated properly.

4.2 Recall and recognition

One of the more helpful distinctions to make about long term memory is the difference between recall and recognition. If you have ever thought: "I can't remember every detail what I'm searching for but I'd know it if I saw it", you are dealing with the distinction between recall and recognition. Another example may help show the difference between the two. If a class of students are shown a 3x3 grid at the start of a lecture on short term memory and then asked to recreate that grid an hour later, this will be a difficult task (with an exception for that one fellow in the back) but if they are shown three different examples where one is the first image then they can quickly identify which image was shown at the beginning. How are they able to do this when they can't really remember what the first one looks like? The answer is that recall is weak but recognition is strong.

We are able to recognize objects that we cannot recall exactly but what kind of factors affect this recall? This is where the idea of distracters is introduced. For further information on the subject, see [3]. Imagine that you are showing someone a series of images and later showing them another image and asking if they recognize it. Obviously their ability to recognize an image depends on how many images were shown and what kind of images was displayed. If the subject was shown, a series of very similar images such as grids that only different in a very small way, the accuracy of recognition goes down. Similarly, if the subject is shown a unique image then the ability to recognize that image again is increased. Such is as we might expect but there are some things about recognition that are unexpected.

One might think that an object that we see every day and have a great deal of experience with should be easily recognizable. Consider for example, the humble penny. It is a common coin and we handle it a great deal in our day to day activities. Yet without looking at a penny, try to identify the correct image in this picture.



This is a more difficult task that you might think, even when we are dealing with an object with which we have had plenty of experience. Nickerson and Adams [4] performed this experiment with pennies and found that only 43% of the subjects were able to correctly recognize an object that they had seen thousands of times before. Their results proved that other factors affected recognition and that the attention given to an object affects the level of recognition.

4.3 Memomonic aids and unusual forms of memory

So far we have discussed visual memory, but the actual processes that make it up are still a mystery. In some ways, we will never be able to completely understand it as it is unlikely that it functions according to a known model. Many theories attempt to describe the mind as if it was some kind of complex computer with distinct parts and code executing in the background. It is not strange that this view of the mind came into popularity in the 70s and 80s when computers were the physically complex objects known. But aside from complexity, computers and minds are objects with completely different properties, origins and compositions. This does not mean that models are useless but we should remember that models are only a tool for understanding memory. We can receive insight into memory through examining the kinds of interactions people have with their memory and unusual forms of memory that some people have.

The process of memory aids shows promise of giving us insight into how our memory works. If we understand the techniques that people use in order to remember objects, perhaps we can understand more about memory in general and visual memory in particular. One classical technique for remembering a list of objects is to associate them with a series of images. This technique dates back to Roman times when rhetoric and oratory were heavily studied and it was necessary for the orator to be able to remember the many topics of his speech and go through them in the correct order. It also helps to remember any arbitrary list of time like a list of U.S presidents for example. The technique might work like this: The first president was George Washington so the first image is a washing machine. The second president is John Adams so the next image is opening washing machine and seeing little atoms floating around. The second image reinforces the first and also establishes the sequence which makes it unlikely that the Washington will be remembered after Adams as that would throw the image sequence out of order. From personal experience, this techniques works better than might be expected and can be a great aid to memory. The question remains though why does the visual image reinforce the arbitrarily word list?

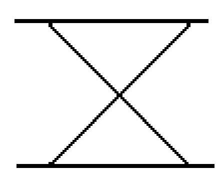
One explanation of the success of this technique is the dual coding model. Bearing in mind the restrictions of models in general, the model helps us understand why this might work. The theory goes that verbal\linguistic information is stored separately from visual/imagistic information. Thus the above technique works because it is stored in two different places and if there are gaps in one area, they can be filled in with the other area. If we forget that Adams comes after Washington, we can remember the images and continue from there. However there is a problem facing this theory. Research on the effect of verbal descriptions of faces [5] shows that applying a verbal description to a face does not aid in our ability to recognize that face. In fact, applying a verbal description impedes our ability to recognize that face again. This is the opposite of the result predicted by the dual coding theory and shows that there are some elements about our memory that are still unknown.

Another area where information can be gleaned about memory is the area of people with unusual forms of memory. Consider the rare case of those with photographic memory. Those with photographic memory claim to be able to view mental images with complete clarity and accuracy. This is a rare ability but it can show us some interesting qualities of our memory. The clear picture-like mental images suggests that our mental images may be similar. This suggests that even though it is difficult for us to perform the same feats of recall that our visual memory works the same.

In addition to those who exhibit unusually enhanced forms of visual memory, there are also those who suffer from impairments to their visual memory. Amnesia or the loss of memory is the common and easily thought of example. Books and movies have popularized the concept of amnesia especially the case of retrograde amnesia where the subject loses memories which occurred prior to the onset of amnesia. The other type of amnesia, anterograde amnesia where the subject loses the ability to generate new memories is also interesting how it shows the separation between long term and short term memory. One of more striking examples of this is the case recorded by Milner in 1966. The subject could normally recall events that occurred in the past and his short term memory enabled him to carry on a normal conversation while the researcher was in the room. However, if the research left the room and returned several minutes later, the subject lost all memory of the conversation. This lends support to the idea that the short term memory and long term memory are different systems.

4.4 Verbal distortion

Our visual memories are not static. Even after an event has generated a memory, that memory can change based on later events and may end up quite different from the original stimuli. This is why asking a witness in court "When did you hear the gunshot?" is a bad question. It is bad because even if the sound that the witness remembers hearing is not a gunshot, describing it as a gunshot will cause the sound to seem more and more gunshot like until it is clearly a gunshot. The same principle applies to our visual memories. If an ambiguous image is shown and a verbal label applied then image will be remembered as being more like the verbal label. The picture shown here can be described as either a table or an hourglass.



If the above image is described as a table, people will remember it as being more table-like than it actually was. If it described as an hourglass, then they will remember it as a figure closer to an hourglass. If a video of two cars colliding is shown and then described as two cars "smashing together" as opposed to "hitting each other", the memory will show the cars moving faster when described as "smashing together". The car experiment was carried out by Loftus [6] but there is debate over exactly what is occurring. Are the original visual memories being altered or are new alternate visual memories being created? It is commonly accepted that visual memory distortion effects exists but how they work is a bit more of a mystery. Loftus and her colleagues hold that the misinformation is actually altering the original images and that the misinformation is irreversibly integrated into the image. But there are others that hold that the misinformation creates a new alternate memory that exists alongside the original image. McCloskey and Zaragoza hold that Lofus incorrectly interpreted the results of her study and that the misinformation does not affect the original image at all. Regardless of which side is correct, both agree that verbal misinformation can affect the ability accurately recall information from the visual memory which raises questions about the nature of our visual memory.

4.5 Analog and Propositional debate

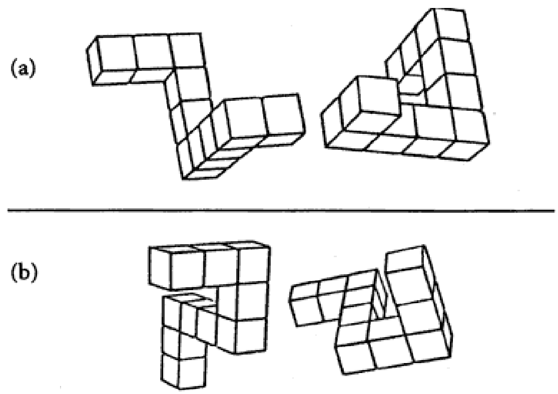
In the sections above, we have examined our visual memory in a roundabout way. We have tried to see what our visual memory is like based on various unusual situations. We have seen that recall and recognition have many factors affecting them, we have seen that memomonic techniques and verbal cues can affect our visual memory. We have been examining the edges of our visual memory and now we are ready to move to the center question. What is our visual memory in truth?

There are two schools of thought on visual memory. One holds that our visual memory is made up of picture-like visual images. These images are distinctly spacial and any operation that can be performed on images can also be performed on our mental images. The idea is that our visual memories are best described a collection of pictures in our heads that we can recall and manipulate at will. This position is supported by our experience with those who display photographic memory who seem to be easily able to do this but there are several objections to this position. The first is that image retrieval is difficult to understand if visual memory consists of pictures stored in our heads. If there is no metadata stored for each image, then how do we summon up an image at will without searching through all of our memory? The second objection is flows from the first; visual images seem to always be interpreted when they are stored in memory yet this does not seem possible under the analog model. The third objection is that the visual memory is often indeterminate. We can summon up a image from our memory but we may not be able to answer questions about the details of that picture.

These objections led Pylshyn to challenge the analog position and suggest that mental images were best thought of as symbolic expressions. This position became known as the propositional position as it held that mental images were constructed of abstract propositions. Instead of holding that mental images were picture-like, the propositional position held that they were constructed from basic language-like concepts and then assembled into images. The end result is a image but the visual memory is encoded in a non-image format.

The merits of these two positions were heavily debated and only experimental evidence would be able to resolve the debate.

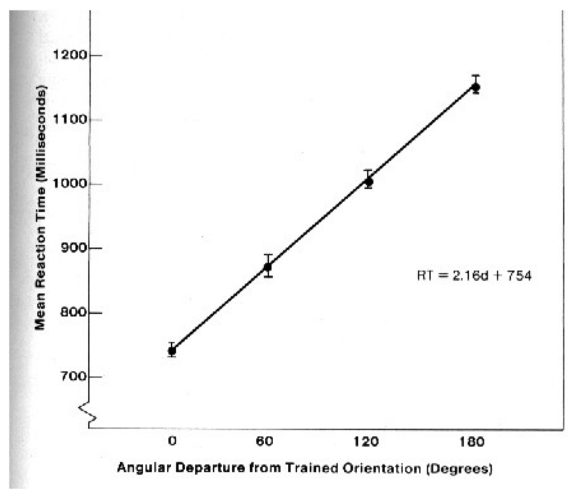
4.6 Mental Rotation and Image Inspection



Mental Rotation Test—Are these two figures the same except for their orientation?

One of the first experiments designed to shed light on the analog and propositional debate was mental rotation. Given the following image, the subject was asked if the two images were the same image related by rotation. The time the subjects spent deciding was measured and the results collected. Now if the analog position holds true, you would expect the subjects to take longer when shown two images with a large rotation and take less time when shown an two images that differed only through a small rotation. The reason being is that the analog position holds that people perform image rotations incrementally or that they gradually rotate an image until it lines up with the other one. They perform visual operations on the picture in their head until it matches the target picture. We would expect that larger rotations will take longer as the person will have to perform more operations in order to align the two images. If the propositional view holds true, we would expect small and large rotations to take about the same time. If instead of pictures, our visual memory consists of propositional statements than the image on the right should be encoded in a manner similar to the following: 3 square column connected by a 90 degree angle to another 3 square column connected by a different 90 degree angle to a 3 square column connected to a single square. If our visual memory consists of information like this, we should be able to look at both images and classify them as same or different without having to rotate them. Thus it should not matter to the response time how big the rotation is between the two images.

The result is shocking. It is not often that experiments produce conclusive results and rarer still when they happen to exactly agree with one of our theories but if you observe the chart of response times for this experiment, you will notice that is exactly the results predicted by the analog position.



The response time is linear in the degree of rotation. This heavily suggests that people rotate the image incrementally until both images have the same rotation and it is possible to tell if they are the same image or not. If the propositional view was correct, the rotation of the image would be a single data point in image information and it should not take any more time to change the rotation from 10 degrees to 170 degrees than it would to change the rotation from 10 degrees to 25 degrees.

These experiments are supported by other experiments. If two images are shown that differ in size, the time required to determine if they are the same image depends on the size difference. In addition to rotation, it appears to scaling is also a linear operation. In a series of image scanning experiments, the hypothesis was that if the analog position holds true, then the amount of time it takes to scan a image should be linear as well. If mental images are picture-like then when they are viewed with the mind's eye, the time it takes to inspect the image will vary systematically with the analog properties of the image being inspected. in the image scanning experiments, the subjects learned a map and studied it until they could reproduce it accurately. This ensured that they had an accurate mental image stored in their memory. They were then asked to span along the image starting from a certain point and ending when they reached another point. It was found that the time spent scanning varied directly on the length of the scanning path and this provided more support for the position that mental images are fundamentally picture like.

4.7 Mental Psychophysics

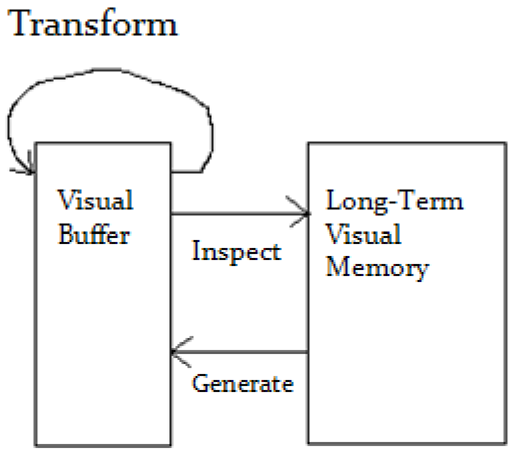
The above experiments suggest that we view our mental images like pictures through something called the mind's eye. It is as if we viewed our mental images through a camera that we could move around the image in order to view it at different angles or pan across the image to scan across. What can we determine about this mind's eye? It turns out that we can determine the viewing angle of the mind's eye camera as well as several other properties from a simple experiment. Several subjects were asked to visualize object and then mentally zoom in until the object overflowed their mental view. At this point, they estimated the distance they were from the object and from this information we can calculate the mental field of vision. It turns out that the mental field of visiois quite limited compared to our eyes' field of vision. While we perceive around 180 degrees in our normal vision, we only perceive 20-40 degrees in our mental vision and we lose the ability to capture detail beyond this narrow field of vision.

4.8 Image reinterpretation

The analog account of visual images has so far been heavily supported by the experimental results but there are some experiments which provide support for the propositional account. One of the positions of the propositional account is that images are stored with an interpretation attached to them. This interpretation is part of the image and cannot be separated from it. There is some basis for believe this to be correct as ambiguous mental images are difficult to reinterpret. To study this, subjects learned a given ambiguous image such as the table/hourglass shown above or the Necker cube. They were told that it was one form and learned the image well enough so that they could reproduce it accurately. They found that the subjects could not reinterpret the image while it was in their heads, but if they drew it out they could now see the other form of the ambiguous image. This lends support to the idea that mental images are encoded with some kind of interpretation in the visual memory as the interpretation remained fixed in their mental image but could be flexible when they were directly perceiving the image.

4.9 Models of Imagery

The debate between the analog and propositional viewpoints continues and is unlikely to be resolved in the sense that one view is proven to be completely right and the other completely wrong. There has been a great deal of experimental evidence on the side of the analog position but there is also reason to believe that propositional viewpoint has some truth to it. Although it appears that mental images have many picture-like qualities, many concepts such as visual memory retrieval and image interpretation are best explained from a propositional viewpoint. These two views both provide valuable insights into our perceptions and thus when we construct a model of imagery, we need to take both approaches into account.



Kosslyn's model of imagery adopts a hybrid approach. Although Kosslyn was a proponent of the analog position, the model he constructed takes elements from each to construct a overarching theory of imagery. It also integrates the ideas presented earlier in this paper on short term memory and combines them with the theory of long term memory. The model has three simple parts: The visual buffer, the long term memory and the operations that interact with the first two. The visual buffer stores input received from the eye and the visual images there are short term analog representations that are constantly being updated, discarded or sent to the long term storage. The long term visual memory stores the kind of visual information that we have been discussing for the last half of this paper. In the Kosslyn model, images are described in propositional terms but each part of the image is linked to an analog representation of the part. The memory of a chair is broken down into separate components but each component has an associated image which can be examined. The operations which form the final part of the model are a set of routines that modify the information stored in either the visual buffer or the long term memory. The transform operation alters the image in the visual buffer, the inspect operation stores an visual buffer image in the long term memory and the generate operation recalls a image from the long term memory storage. To see how this model works, consider the case of image scanning. The subject has learned the image so it is stored in his long term memory. The generate operation moves the image from the long term memory to the visual buffer and then the transform operation is applied and a scanning operations begins on the image and when the scanning is complete, the subject indicates that he is finished.

4.10 Visual imagery and Perception

The last topic that we will discuss is the connection between visual perception and visual imagery. In this paper, we have primarily focused on visual imagery, i.e. the images that memory forms. In our discussion of the mind's eye camera and the other qualities of visual imagery, you might have noticed that we discussed them in very similar terms to our visual perception. The mind's eye was described as a form of camera that perceived an image and other perception terms were used. This implies that there may be a link between our memory and our perception. There is good experimental evidence for this being the case. If subjects are asked to perform a demanding visual imagery task such constructing a detailed image from memory, their ability to perceive visual stimulus decreased. In the same way, once they focused on perceiving, the time it took from the subjects to perform feats of mental construction increased.

There is also neurological support for a link between visual imagery and perception. In brain damaged subjects, the loss of perception is often correlated with the loss of some memory ability. A subject who had difficulty describing the details of objects also had difficulties remembering the minor details of objects that he had seen. A subject who had difficulty perceiving temporal involvement also had difficulty remembering the order of a series of images. In addition, brain activity scans of normal subjects showed similar levels and areas of activity when the subjects were engaging in both acts of perception and acts of visual imagery. Just how much overlap there is between perception and memory is still a mystery but it appears that there is a strong connection between the two.

5. Conclusion

There is no doubt that memory plays a vital role in our lives. Our visual memory allows us to unify our visual perceptions into one whole and function in a constantly changing world. In this paper, we have seen how visual memory takes all different kinds of forms from the icon memory which allows us to remember events that happen incredibly quickly, to our long term memory which can summon up the farther past. We have seen how all the parts of memory, both short term memory and long term memory, work together to accomplish complex and amazing tasks. Truly, the importance of memory is worth remembering.

6. References

[1] Biederman, I. "Recognition-by-components: a theory of human image understanding." *Psychol Rev*. 1987.

[2] Ullman, S. "Visual routines." *Cognition,* 1984.

[3] Hines, D. "Recognition of random shapes followed at varying delays by attended or unattended shapes, digits, and line grids." *Human Learning,* 1973.

[4] Nickerson, R.S., and Adams, J.J. "Long-term memory for a common object." *Cognitive Psychology*, 11, 287-307. (1979)

[5] Dodson, C. Johnson, M. Schooler, J."The verbal overshadowing effect:Why descriptions impair face recognition" *Memory & Cognition* 25 (2), 129-139, 1997.

[6] Loftus, E.F., & Palmer, J.C. "Reconstruction of automobile destruction: An example of the interaction between language and memory". *Journal of Verbal Learning and Verbal Behavior, 13,* 585-589. 1974.