The interface between priming memory and the working memory

Graeme E. Smith, GreySmith Institute of Advanced Studies
Edmonton, Alberta, Canada
Abstract:

There is an interface between Priming Memory and Working Memory that acts to limit the amount of data in Working Memory, and allows Priming Memory to be actively maintained so that the memories stay intact while they are in Working Memory. This interface converts from functional clusters to chunks automatically as part of the transfer from priming to working memory. As a result, memories in working memory can be rehearsed in order to actively maintain them. This represents more than a change of state, it represents a radical change in the way that the memories are accessed. Priming memory without this stage can only be extended it cannot be refreshed and so there is a limit of about 3 seconds to how long it will last before it begins to degrade. With priming, an actively maintained and refreshed signal can last almost indefinitely.

Keywords:

Priming Memory, Working Memory, Memory Interface, Delay Memory, Rehearsal, Refresh, Active Maintenance, functional Cluster, Similarity Selection, Chunking

Priming Memory is implicit, and if Marr [Marr D., 1970] is to be believed Content Accessible memory. As such it has some rather strict limitations on how it can be used. It is created by the formation of Functional Clusters [Edelman G. 1990] through a process I call Similarity Selection. The most striking aspect of this, is that priming memory cannot be retrieved except by stimulation with a similar stimulus to one already stored.

Working Memory is often explicit, and can be actively maintained. The difference is striking, and can be mostly put down to the Interface between priming memory and working memory, a second type of implicit memory called delay memory. Essentially Delay memory changes the interface between the brain and priming memory so that it becomes retrievable in chunks. How exactly this conversion happens is still somewhat speculative, but we know that it must happen because Working Memory is actively maintained and accessible by chunks.

Since Millers work in the 1940’s [Miller G., 1955] we have known that the storage size of Working Memory is quite small. Millers own estimate was 7 ± 2 but more recent estimates have dropped the value to 3 or 4. However, there have also been span tests that show chinese students to have 11 items in their digit span which leads me to the conclusion that the results should be 7 ± 4. The number of items is limited, but depends on arrangement which is why chunking was invented to explain why some people could store more than others by rearranging the items into groups or chunks.

Information theory recognizes two types of memory, Content accessible, and Place-code addressable. Chunks are lists of place-codes which is problematic for neuroscientists because this means that neural networks have to be able to be place-coded. Nothing I have seen so far indicates that connectionists have found an easy
way to place-code neural networks. My work is not an easy way to place-code, but it suggests that delay memory could be place-coded with a relatively simple interface, involving a serial dependency.

The theory is that the superficial laminae are the same across allocortical and isocortical tissues, and represent priming memory, but the deep laminae laminae IV, V, and VI present an interface that can be, with inclusion of other brain functions, equivalent to a place-coding addressing scheme. Laminae IV doesn't contribute much to the place-code addressing scheme, it is just the granular layer where thalamus inputs from sensory areas are presented to the isocortical column.

It is characterized by highly specific fibers. Laminae V is also connected to the thalamus, but by what are called non-specific fibers that terminate on as many as thousands of neurons. This is the projected interface layer between the place-code addressing scheme and the isocortical column.

Laminae VI on the other hand is mostly to do with the isocortical column, and it's integrity, it causes the column to act as if it were a single neuron and fire as a unit. The idea is that the non-specific fibers pre-activate laminae V neurons which act like mini-columns and feed the column as a whole. The non-specific fibers are degenerate in their signaling causing quite a bit of redundancy and overlap, pre-activation of multiple mini-columns is equivalent to addressing a sub-set of those mini-columns and activating a column where the mini-columns are shared within the same column.

While this is enough to implement a place-code, it is not enough to implement chunking. Instead a list of active columns is needed to implement a chunk. While the actual interface that allows this has not been discovered, it is thought be some that the interface might lie in the Ventro-Lateral Prefrontal cortex. In an area called area 46. Approximately 1/3 of this area is taken up with a buffer which echoes cortex areas that are in the focus of attention. It was first discovered as proof of numeracy in monkeys [Neider A., Merten K., 2007], and later as somewhere that showed colors in humans. When it was realized that the area responded to whatever was in the attention at the time, and it's role as a buffer became understood it was thought to be Working Memory, but that was soon questioned as well [Passingham R., Wise S., 2012] because it is in the prefrontal cortex which murine mammals do not have, but what murine mammals do have, is a demonstrable working memory. While the last word has not yet been seen on this topic, it does show that the area is actively involved with buffering different areas of the cortex. Which makes it a candidate for a specialized buffer for primate working memory, and a possible location for a mechanism that converts to chunks.

That leaves the serial dependency that limits the size of working memory to understand. Near as I can tell, the serial dependency is created because only one chunk can be analyzed at a time. This means that the analysis step in the conversion from Functional Cluster to Chunk, is rate limiting and acts as a brake on communication between the priming memory and working memory. This serial
dependency is likely caused by the fact that the memory has to isolate one buffer full of data before it can be converted. As a result only one chunk can be converted at a time.

Conclusions:

There is an interface between priming memory and working memory that acts to limit the size of working memory, and convert it from content accessible to place-code addressable chunks. It is likely that area 46 is involved with this task. I call this interface Delay Memory. And trace it to the deep laminae of the isocortical tissues.

References:


Cabeza R. (1994) A dissociation between two implicit conceptual tests supports the distinction between types of processing Psychonomic Bulletin and Review 1(4):505-508


Smith G., (2013) *Primary memory is priming memory: seen through the working memory viewpoint* peerj preprints doi:10.7287/peerj.preprints.36v1
