European Research Council

ERC Advanced Grant 2011
Research proposal (Part B1)

Memory Mechanisms in Man and Machine

M4

Cover Page:
- Name of the Principal Investigator (PI) : Simon J Thorpe
- Name of the PI's host institution for the project : CNRS
- Proposal duration in months : 60 months

Proposal summary

The project aims to validate a set of 10 provocative claims.

1) Humans can recognise visual and auditory stimuli that they have not experienced for decades.
2) Recognition after very long delays is possible without ever reactivating the memory trace in the intervening period.
3) These very long term memories require an initial memorisation phase, during which memory strength increases roughly linearly with the number of presentations
4) A few tens of presentations can be enough to form a memory that can last a lifetime.
5) Attention-related oscillatory brain activity can help store memories efficiently and rapidly
6) Storing such very long-term memories involves the creation of highly selective "Grandmother Cells" that only fire if the original training stimulus is experienced again.
7) The neocortex contains large numbers of totally silent cells ("Neocortical Dark Matter") that constitute the long-term memory store.
8) Grandmother Cells can be produced using simple spiking neural network models with Spike-Time Dependent Plasticity (STDP) and competitive inhibitory lateral connections.
9) This selectivity only requires binary synaptic weights that are either "on" or "off", greatly simplifying the problem of maintaining the memory over long periods.
10) Artificial systems using memristor-like devices can implement the same principles, allowing the development of powerful new processing architectures that could replace conventional computing hardware.

We will test these radical claims with a highly interdisciplinary approach involving psychology, neuroscience, computational modeling and hardware development. Novel experimental paradigms will study the formation and maintenance of very long term sensory memories. They will be combined with imaging techniques including fMRI imaging, EEG recording, and intracerebral recording from epileptic patients. In parallel, computer simulations using networks of spiking neurons equipped with Spike-Time Dependent Plasticity will attempt to explain the experimental results, but also used to develop bio-inspired hardware systems that can reproduce the learning capacity of the brain using novel memristor-based technologies.

Section 1c: Extended Synopsis of the project proposal

Introduction

This proposal is a completely revised version of a project that I submitted to the ERC Advanced Grant program in 2011 and which was shortlisted but narrowly failed to make the final selection. I have done my best to answer the points raised by the committee, and have updated the project in the light of some recent pilot work that we have been doing.

The proposal is a radical departure from the work on ultra-rapid visual processing for which my group is well known. The core of the project is a set of ten highly provocative and (probably) controversial claims about human sensory memories.

1) Humans can recognise visual and auditory stimuli that they have not experienced for decades.
2) Recognition after very long delays is possible without ever reactivating the memory trace in the intervening period
3) These very long term memories require an initial memorisation phase, during which memory strength increases roughly linearly with the number of presentations
4) A few tens of presentations can be enough to form a memory that can last a lifetime.
5) Attention-related oscillatory brain activity can help store memories efficiently and rapidly
6) Storing such very long-term memories involves the creation of highly selective "Grandmother Cells" that only fire if the original training stimulus is experienced again.
7) The neocortex contains large numbers of totally silent cells ("Neocortical Dark Matter") that constitute the long-term memory store.
8) Grandmother Cells can be produced using simple spiking neural network models with Spike-Time Dependent Plasticity (STDP) and competitive inhibitory lateral connections.
9) This selectivity only requires binary synaptic weights that are either "on" or "off", greatly simplifying the problem of maintaining the memory over long periods.
10) Artificial systems using memristor-like devices can implement the same principles, allowing the development of powerful new processing architectures that could replace conventional computing hardware.

Validating these ten claims will clearly be a major challenge. However, we already have a number of preliminary results that demonstrate that the ideas presented here are perfectly plausible. During the five years of the project our aim will be to test the validity of all ten propositions using a highly interdisciplinary approach that combines novel protocols in experimental psychology, neuroscience, computer modelling and hardware development.

The Ten Claims

1. Humans can recognise visual and auditory stimuli that they have not experienced for decades.

When asked, most adults are convinced that they can recognise visual and auditory stimuli that they have not experienced for decades. For example, people born in the 1950s may be able to recognise the theme from an old television program (e.g. "The Lone Ranger") that they watched as children in the early 1960s. Surprisingly, there has been remarkably little research into these sorts of very long term memories. One well-known study demonstrated that people could still recognise photographs of classmates from college yearbooks over periods of up to 50 years (Bahrick et al 1975), but of course there was no control over the sorts of experience that led to the formation of these long-term memories. One of the novel approaches that we will develop will test people's ability to name old radio and TV themes from the 1950s and 1960s. By using information provided by the INA (Institut National de l'Audiovisuel) in France, and the BBC in the UK, we should be able to derive functions describing the probability that such extremely long-term memories are formed as a function of the number of times that the program was seen or heard originally. This goes well beyond any previous studies of these very long-term memories.

2. Recognition after very long delays is possible without ever reactivating the memory trace

Even if we succeed in demonstrating that people really can recognise radio and TV themes from the 1950s and 1960s, this would not prove that the memory trace can survive without being reactivated in the intervening period. If we were lucky, we might be able to find audio and video clips that we are sure have never been rebroadcast for several decades, and which therefore could not have been seen or heard since. But even then we cannot rule out the possibility that subjects may have (even unwittingly) thought or even dreamt about the program at some point. However, in a recent study, we were able to show that our brains can form robust sensory memories for meaningless auditory noise patterns, and that these memories can last for 2-3 weeks at least (Agus et al 2010). Given the nature of the stimuli, we can effectively rule out any possibility that the subjects could reactivate the stimulus "in their heads". This therefore seems to be a clear case where (relatively) long-term sensory memories can be maintained without reactivation. One of the key aims of the current project will be to see whether such memories can remain intact for longer periods up to several years.

3. During the initial memorization phase, the strength of sensory memories increase roughly linearly with the number of presentations

Another of the striking results in the Agus et al study is that the strength of the memory appeared to increase almost linearly with the number of presentations. In other words, memory strength with 10 presentations was roughly twice that seen with 5 presentations. Here again, one of the key objectives of the project will be to see whether this rule of thumb applies to more typical material, including snatches of music, images, and video clips.

4. A few tens of presentations can be enough to form a memory that can last a lifetime

This is clearly very tentative. However, our auditory noise learning data shows clearly that even with just 30-40 presentations, our subjects had formed robust memories that could last for weeks. It is not inconceivable that these results could be extended to much longer periods. For example, my personal experience suggests that the TV shows that I can recognise easily from my childhood in the 1960s are the ones that where I was a regular viewer and watched every episode. It is here that the web-based analysis of people's ability to recognise old TV and radio themes proposed in this project could be very illuminating. Specifically, this work will attempt to determine if there is a systematic function that links the number of exposures and the probability of forming a robust memory trace.

5. Attention-related oscillatory brain activity can help store memories efficiently and rapidly

This is another educated guess that needs to be backed up by hard experimental testing. We know that humans can memorize complex visual stimuli even when they have just a few seconds to inspect the image. This was originally shown over four decades ago (Shepard 1967; Standing 1973), but there has been a fascinating series of recent studies from Aude Oliva's group at MIT showing just how detailed these sensory memories can be (Brady et al 2008; Konkle et al 2010a; b). In those experiments, the subjects were typically tested just a few hours later or the next day, so it is not yet clear that we are talking about really long-term memories, although Mandler and Richey (1977) did demonstrate certain forms of memory for line drawings extending to 4 months. This is something that we will specifically test. Nevertheless, if long-term sensory memories can indeed be formed with just 3-5 seconds of exposure, this might seem to be at odds with my suggestion that a few tens of presentations are required. However, our modelling work (see later) has shown how oscillatory activity can help memory formation (Masquelier et al 2009b), and there is a wealth of evidence showing that oscillatory activity in the alpha-band (8-13 Hz) plays a role in attention (Jensen et al 2012; Klimesch et al 2011; Mo et al 2011). Could it be that oscillatory activity in the visual system effectively allows several tens of presentations to be performed internally even when the stimulus is only visible for a few seconds? This will be tested experimentally during the project by seeing whether there is a correlation between the degree of alpha activation and the ability to recognise a visual image presented for just a few seconds.

6. Storing very long-term memories involves the creation of highly selective "Grandmother Cells"

This claim is undoubtedly the most controversial one. The idea that we may have highly selective neurons that will only respond to very specific stimuli has never really been popular among neuroscientists – at least in public. Horace Barlow's suggestion that "Perception corresponds to the activity of a small selection from the very numerous high-level neurons, each of which corresponds to a pattern of external events of the order of complexity of a word", remains no more than a hypothesis

(Barlow 1972). Only a few authors have seriously argued in favour of such a proposal (see for example, (Bowers 2009; 2011)), and such views have been criticised (Plaut & McClelland 2010; Quian Quiroga & Kreiman 2010). There have been a series of fascinating studies showing that single neurons in the human medial temporal lobe can respond selectively to particular individuals and places such as "Jennifer Anniston", "Sydney Opera House", "Bill Clinton" (Mukamel & Fried 2012; Quian Quiroga et al 2009; Quian Quiroga et al 2005). However, even the authors of these studies argue that such results should not be considered as evidence for Grandmother Cell coding (Quian Quiroga et al 2008). It is certainly true that the hit-rate for finding such cells is way too high. When tested with a set of images, the chances that a neuron in the human hippocampus will respond to any particular stimulus is around 0.5% (Waydo et al 2006) – with true Grandmother Cell coding, activation should be much sparser.

Nevertheless, I have repeatedly argued that there are no strong arguments for ruling out "grandmother cell" coding (Thorpe 1995; Thorpe 2002), and recently, I have argued that Grandmother Cells may be the only way to explain how the brain can store memories intact for decades without the need for reactivation (Thorpe 2011). The key mechanism underlying the hypothesis is the concept of Spike-Time Dependent Plasticity (STDP), the idea that synaptic weights are effectively only modified when a neuron fires. Specifically, any incoming synapses that fire just before the neuron fires are strengthened, whereas inputs that fire after the neuron fires get weakened. Such learning rules, which are effectively a specific implementation of Donald Hebb's classic hypothesis (Hebb 1949), were first demonstrated in the late 1990s (Bi & Poo 2001; Markram et al 1997), and have since been studied extensively. One consequence of such rules is that if a neuron so elective that it has no spontaneous activity, it follows that, in the absence of its preferred stimulus it could perfectly well remain totally silent for years or even decades. Under such conditions, it would never spike, and so its pattern of synaptic weights would never modify, allowing the memory to remain intact indefinitely. To the best of my knowledge this hypothesis is totally original and constitutes the only currently available hypothesis consistent with the data.

My claim is that conventional distributed representational codes would be incapable of retaining sensory memories for such long periods. While such networks can certainly store a limited number of patterns in a reasonably robust way (Baum et al 1988; Hopfield 1982), any tendency of neurons to fire to a range of different stimuli will invariably lead to the memory trace being overwritten at some point.

7. The neocortex contains large numbers of totally silent cells ("Neocortical Dark Matter")

While this claim may sound surprising, it is actually one for which there is a considerable amount of support, as pointed out in a recent review paper (Shoham et al 2006). As long ago as 1968, David Robinson had already noted that when neurophysiologists record from areas such as the visual cortex, they should normally be able to record from about 100 neurons per track. In fact, in a typical recording session, most neurophysiologists consider themselves lucky if they can record more than a handful of cells, and Robinson suggested that one explanation could be that a large percentage of the neurons could be silent, and thus invisible to the neurophysiologists electrode (Robinson 1968). One way to force silent neurons into life even when they have no spontaneous activity is to activate them antidromically by electrically stimulating their output fibres. When he did this, Swadlow reported that around 50% of the neurons in rabbit V1 projecting to other brain structures have spontaneous firing rates below 0.1 Hz! (Swadlow 1988). In recent years, it has been possible to use new techniques such as two-photon microscopy to investigate neural activity across populations of neurons simultaneously (Katona et al 2012; Ohki et al 2005), or specialised electrode techniques that can record from many sites simultaneously (Blanche et al 2005). However, while these methods have enormous potential, we do not yet have a clear picture about the precise proportion of neurons that are silent. For this reason, my suggestion that a substantial proportion of completely silent cortical neurons could be the neural substrate of very long term memories remains perfectly plausible.

8. The formation of Grandmother Cells can be explained by a very simple model

My claim that Grandmother Cells could underly very long term sensory memories requires a plausible mechanism to allow such highly selective responses to be produced. Intriguingly, the question of how to generate highly selective responses has received surprisingly little direct attention in the literature, probably because many researchers reject the notion out of hand. However, our own recent modelling work has demonstrated that, in fact, highly selective neuronal responses can be produced using a remarkably simple mechanism. Specifically, we have been using models based on very simple leaky

integrate-and-fire neurons equipped with a basic Spike-Time Dependent Plasticity rule. Several years ago we demonstrated that when such neurons are presented with repeating patterns of inputs, high synaptic weights will systematically concentrate on the early firing inputs (Guyonneau et al 2005). We went on to show that because of this, neurons will naturally learn to become selective to patterns of inputs that occur repeatedly (Masquelier et al 2008; Masquelier & Thorpe 2007). In other studies, we demonstrated that when multiple neurons are listening to the same set of inputs, and if there are inhibitory connections between them, they form a competitive learning mechanism in which each neuron will learn to respond to a different input pattern (Masquelier et al 2009a). But the most convincing data comes from recent work using a modified STDP rule that allows neurons to become very selective to any repeating stimuli (Bichler et al 2011, 2012).

9. This selectivity only requires binary synaptic weights

One particularly striking feature of our STDP based learning model is that once the learning is complete, the synaptic weights are either fully on, or fully off. This means that in order to keep a memory intact, it is enough simply to keep all the synapses that work in the fully on state. The other synapses are no longer needed. This makes it much easier for the brain to keep the memories intact for long periods, because it is not necessary to store fractional weight strengths. Any alternative proposal that required intermediate weight strengths to be stored accurately over time would face major difficulties in explaining very long-term sensory memories.

10. Artificial systems using memristor-like devices can implement the same principles

There are currently a number of research programs looking at the potential of nanotechnology devices called "memristors" which are semiconductor junctions whose resistance can be modified by applying a voltage above a certain critical threshold (Strukov et al 2008). This feature allows them to be used to store memories, and it has recently been demonstrated that such devices can, in principal, be used to implement a form of STDP-like learning (Perez-Carrasco et al 2010; Rachmuth et al 2011; Zamarreno-Ramos et al 2011). This opens a path for developing completely revolutionary computing architectures that can directly implement the sorts of processing schemas using by biological systems. My student, Olivier Bichler, has already demonstrated that it is theoretically possible to implement circuits that can learn to recognise particular visual patterns in this way (Bichler et al 2011; 2012). Together with his colleague Damien Querlioz, he has also demonstrated that such devices would have a truly remarkable level of tolerance to component defects (Querlioz et al 2011). While these are early days, such results are extremely encouraging because they point towards a way of implementing biological inspired mechanisms for learning about complex sensory stimuli in hardware. If the project succeeds, it will be possible to use insights from biological learning mechanisms to develop revolutionary information processing architectures that could replace conventional computing systems.

The Overall Workplan

These 10 claims will be tested during the course of this five-year project using a highly interdisciplinary approach that combines techniques from experimental psychology with neuroscience, computational modelling and hardware development.

First, my claims about the ability of humans to store sensory memories that can be maintained for several decades without the need for reactivation will be tested using a range of novel experimental approaches. We will build on our recent work that showed that subjects can form robust memories for meaningless auditory noise that can last for at least a few weeks by testing recognition of similar stimuli with longer periods extending to several years. We will continue with pilot studies using fRMI and Event-Related Potentials that have already demonstrated that we can analyse the neural structures involved in the formation of these auditory memories.

Second, we will tackle the question of whether humans really can recognise visual and auditory stimuli that they have not experienced for several decades by testing their ability to recognise old radio and TV themes from the 1950s and 1960s. For this we will make use of archive material from the French INA (Institut National de l'Audiovisuel) and the BBC in the UK for which there is detailed information about the number of times particular programs were shown. Initial experiments will be performed under laboratory conditions, but we plan to use web-based testing methods to investigate this sort of remote memory in thousands of subjects.

Third, we will study the formation of these very long-term sensory memories in the laboratory using procedures that mimic the experience of repeatedly seeing or hearing a radio or TV theme. Specifically, we will ask subjects to associate an arbitrary verbal label with a brief sensory stimulus which can be either just an auditory clip (akin to learning to name a piece of music), a static image, a short video sequence, or the equivalent of a movie clip with both sound and vision. A key question will be to see how the strength of the sensory memory changes as a function of the number of times the stimulus is presented, and the separation between each presentation. We will also use EEG recordings during the training sessions to see whether particular patterns of oscillatory brain activity can predict whether or not particular a given sensory memory can be formed.

Fourth, we will develop large-scale simulations using networks of spiking neurons equipped with Spike-Time Dependent Plasticity (STDP) that will be used to model the formation of both visual and auditory memories using paradigms that replicate our experimental studies. Our initial results suggest that our latest STDP-inspired learning rule can directly explain our recent results on the formation of memories for meaningless auditory noise stimuli, and we will attempt to develop equivalent models for all of the experimental studies.

Finally, we will use the results of these neural network simulation studies to develop novel memristor-based hardware that physically implement the synaptic weight changes associated with STDP. If successful, we will be able to produce an artificial system capable of learning auditory and visual stimuli using exactly the same principles that we believe allow humans to form robust sensory memories that can last for a lifetime.

Impact

The consequences of the project should be highly significant in three main areas. The first area concerns our **scientific understanding** of how the brain functions. We hope to be able to provide strong support for the radical hypothesis that long-term sensory memories require the formation of highly selective "grand-mother cells", cells that will remain totally silent for decades unless the original stimulus is presented again. We should be able to demonstrate how the formation of such highly selective neurons occurs virtually automatically whenever a particular pattern of sensory activation recurs repeatedly. The implication is that the neocortex may contain substantial numbers of totally inactive neurons – a sort of "Neocortical Dark Matter" – normally invisible to the neurophysiologist's electrode. While the current project will concentrate on the formation of memories for sensory stimuli (simply because they are easy to manipulate experimentally), we would argue that exactly the same rules could be apply to the learning based on the pattern of firing in any neural system. The implications for our understanding of brain function could hardly be more profound.

The second area of impact concerns the **development of highly innovative new technologies** that can be used to implement biologically inspired learning mechanisms in artificial systems. If the memristor implementation of STDP-like learning rules is successful, it will be possible to build chips that can implement neurons with tens of thousands of synapses in an extremely compact and energy efficient form. Such systems would also be very fault-tolerant, and could potentially come to replace the conventional Von Neumann processing architectures found in virtually all current computers. Again, it would be difficult to imagine a more profound revolution in technology, and it could clearly have enormous economic implications.

The third area lies in the project's impact on **Society**. If the project succeeds, we will be able to elucidate the keys parameters that determine whether a newly formed memory can last throughout life. Our hypothesis is that the strength of the memory trace increases roughly linearly with the number of times that the stimulus is presented, but that they pattern of presentation is also critical. There may potentially be an optimum way to present the stimuli to have the best chance of the memory lasting. Would 20 presentations on 20 successive days work best? Or would it be better to regroup the training experiments into (say) four days with 5 closely spaced learning experiences? It is surely obvious that this sort of work could be vital to help teachers determine the best way to organize material to obtain optimal results. Importantly, the work presented here is not simply descriptive. Our aim is not simply to say that this particular training sequence works best, but to explain precisely **why** it works best on the basis of the underlying neural mechanisms.

Section 2: The Project proposal

2a. State-of-the-art and objectives

Introduction - Studies of long-term memory

We have probably all had the experience of recognising something we have not seen or heard for a very long time. It might be an old school friend that you meet by chance in the street and who you have not seen for decades. It might be an old picture book that you read as a child and which you discover in a second hand book shop. It might be a record of Danny Kaye singing "The Ugly Duckling" that you used to listen to as a kid and that you hear on the radio. Or it might be the theme to some old TV or Radio program not rebroadcast since childhood that you run in to by chance one day when zapping between channels. When something like this happens, many of us may feel convinced that we can achieve these feats of recognition without ever having thought about the item in the meantime. If so, it would surely have major implications for the way in which memories are stored in the brain. But can you really be sure that the memories have not been reactivated at some time? Maybe you dreamt about them? How could you argue against a scientist who proposes that the function of dreaming is to reactivate old memories in order to refresh them? This last suggestion may seem far-fetched, but is actually very difficult (impossible?) to refute.

Although a great deal is known about human memory (Buckner & Wheeler 2001; Dudai 1997; Moscovitch et al 2006; Tulving 1985; 2002), there is surprisingly little experimental work that has looked at this sort of very long term memory. The classic study is without doubt Bahrick's work on the ability of people to recognise classmates from college year-books over periods of up to 50 years (Bahrick et al 1975, see also Bruck et al 1991). Bahrick also published work on the recall of 2nd languages over 50 years (Bahrick 1984) as well as recall of knowledge about the geography of a city where people studied as students (Bahrick 1983). Such studies have made in possible to establish the forgetting curve – how much you can remember as a function of the delay since the learning phase. For example, knowledge of Spanish learned at school drops off roughly exponentially in the first 3-6 years, but then remains stable for 30 or more years (Bahrick 1984). This is a classic problem in memory research, dating back to Ebbinghaus (1913). However, the shapes of the long-term memory functions are highly variable, and depend enormously on the material (Rubin & Wenzel 1996; White 2001). One reason for this variability may lay in the fact that there is no way of controlling whether or not people actively recall information in the intervening period.

Another active area of research concerns autobiographical memory – our ability to recall information about our own personal history (Conway & Bekerian 1987; Morrison & Conway 2010), and a number of studies have looked at the brain structures that are activated during the activation of such memories (Conway et al 2002; Haist et al 2001; Maguire & Frith 2003; Piolino et al 2004; Viard et al 2007). There is also a substantial literature on so-called "flashbulb" memories for highly emotionally charged events such as the events of September 11th 2001 (Conway et al 2009, Hirst et al 2009). However, while such studies certainly demonstrate that we have memories that can persist for several decades, there are many unanswered and even unasked questions. In particular, no studies appear to have asked whether these long-term memories can be maintained without the need to reactivate them at some point. And even if we suppose that we can have memories that can be maintained for decades without reactivation (as most people probably suspect), what is known about the factors that determine what memories get stored in the first place? The simple answer is that we know remarkably little.

Yet, demonstrating that very long term memories for visual and auditory stimuli can be maintained without the need for reactivation would be a very significant discovery. To understand why, consider the problem of storing a memory trace intact in biological hardware. Most researchers would accept that the basic mechanism underlying the formation of memories involves modifying the strength of the synaptic connections between neurons. For the sorts of memories that we are talking about, it is likely that the long-term memories are stored in neocortex since other structures involved in memory formation such as the hippocampus do not appear to be essential for such remote memories (Bayley et al 2005; Bayley et al 2006). A cortical locus is also suggested by Wilder Penfield's classic research in the 1950s that demonstrated that direct electrical stimulation of the neocortex can result in reliable reactivation of particular memories (Penfield 1958; Penfield & Perot 1963).

Recent anatomical research has demonstrated that the human neocortex contains around 16 billion neurons (Azevedo et al 2009) each of which can potentially connect to a few thousand other neurons, although it is unclear just how many of those connections are functional. Presumably, most of our long-term memories are stored in the pattern of connections between these neocortical neurons (Feldman 2009). Furthermore, it is generally believed that synaptic connection strengths are continuously being modified by ongoing activity (Trachtenberg et al 2002). How then could a particular pattern of connections be maintained over periods of several decades? A further problem is that the molecular building blocks that are used to make synaptic connections are continuously being renewed (Wang et al 2006). Essentially all the structural proteins that are used to build a given synapse are completely renewed every few weeks or months (Bredt & Nicoll 2003). How then can we suppose that the connection strength of such a synapse could be maintained for a long period of time? Indeed, the problem is so significant that a recent paper suggested that the only way to store memories throughout life would be to store the information in DNA molecules (Arshavsky 2006), a move that would effectively throw away virtually all we know about the brain mechanisms of memory.

Grandmother Cells, Neocortical Dark Matter and Very Long-Term Memories

I recently made a radical suggestion that I believe explains how these very long-term memories could be maintained (Thorpe 2011). I proposed that our ability to recognise a particular stimulus after a long delay depends on the formation of highly selective neurons, neurons that are sufficiently selective that they will not fire unless the original stimulus is experienced again. Such neurons, which would effectively be a sort of **"grandmother cell"** (Bowers 2009; Gross 2002), could remain dormant for long periods, potentially decades, allowing the original stimulus to be maintained in memory. Higher-order association areas could contain tens of millions of such neurons, each tuned to particular sensory stimuli, but that would effectively be a sort of **"neocortical dark matter"**, having no spontaneous activity and thus effectively invisible to conventional recording techniques.

The key mechanism underlying the hypothesis is the concept of **Spike-Time Dependent Plasticity (STDP)**, the idea that synaptic weights are effectively only modified when a neuron fires (Markram et al 2011). Such learning rules, which are effectively a specific implementation of Donald Hebb's classic plasticity hypothesis (Hebb 1949) were first demonstrated experimentally in the late 1990s (Bi & Poo 2001; Markram et al 1997), and have since been studied extensively (Caporale & Dan 2008). In the standard STDP version, any incoming synapses that fire just before the neuron fires are strengthened, whereas inputs that fire after the neuron fires get weakened (see Figure 3 for an illustration).

My own group has shown that STDP concentrates high connection strengths on early firing inputs (Guyonneau et al 2005), and that this simple property means that neurons will naturally learn to become selective to patterns of inputs that occur repeatedly (Masquelier & Thorpe 2007). For example, when we presented a network with images from the Caltech Face data base, less than one hundred presentations were enough for neurons in the network to start becoming selective to face features. The neurons become selective to faces simply because those features occurred most frequently in the set of images used for training. Importantly, this learning is completely unsupervised – there is no need for any instructions about what should be learned. More recently, we demonstrated that this STDP based learning is so powerful that it allows neurons to learn to respond to repeating patterns in the firing patterns of thousands of afferent fibres (Masquelier et al 2008). Specifically, a randomly chosen pattern lasting just 50 ms that only involves a small subset of the inputs can be detected after only a few tens of repetitions. And wiithin a few minutes, the neuron can learn to respond at the very start of the pattern. In further studies, we have demonstrated that when multiple neurons are listening to the same set of inputs, and if there are inhibitory connections between them, they form a competitive learning mechanism in which each neuron will learn to respond to different components of the input patterns (Masquelier et al 2009a).

However, in the context of the present discussion, there is an even more important consequence of this form of STDP-based learning rule. If STDP is literally true, a neuron that never fires will never change its connections. This means that a neuron that learns to respond selectively to a given stimulus could potentially keep its pattern of connections indefinitely. And if the original training stimulus is never repeated, the neuron may never fire again. The neuron should still be there even decades later, ready to spring into life if the original stimulus is shown again.

The suggestion that the neocortex may contain extremely selective "grandmother cells" that only respond to very specific stimuli is, to put it mildly, extremely controversial. Although I have argued that there are no strong grounds for ruling out this sort of highly localist coding (Thorpe 1995; Thorpe 2002), few researchers appear to take the idea seriously—with a few exceptions such as Bowers (2009; 2010). Nevertheless, there have been reports of highly sparse coding in both the insect olfactory system (Finelli et al 2008; Jortner et al 2007) and in songbirds (Hahnloser et al 2004). And, in humans, there have been very impressive demonstrations of highly selective single neurons in the medial temporal lobe of human epileptic patients. There are reports of neurons that can fire selectively to essentially any image of a particular individual (for example, Jennifer Anniston, Halle Berry or Bill Clinton). Similar results have been reported for particular places such as the Sydney Opera House, or the Taj Mahal (Quian Quiroga et al 2005). In addition, these authors have shown that a given neuron can respond to the same percept presented via multiple modalities—visual (photograph), text (the name of the object or person), or auditory (the person's name) (Quian Quiroga et al 2009). Furthermore, such neurons tend to respond preferentially to personally relevant images, including members of the patient's close family (Viskontas et al 2009). However, even the authors themselves do not believe that such neurons can be considered to be Grandmother cells (Quian Quiroga et al 2008), essentially because the hit rate (the probability that a given neuron will respond to a particular stimulus) is way too high. Indeed, if the cells were involved in encoding a single percept over the lifetime of the subject, one would expect them to fire extremely rarely. However, neurons in the human medial temporal lobe typically respond to about 0.5% of the stimuli—a value that is far too high for true "grandmother cell" coding.

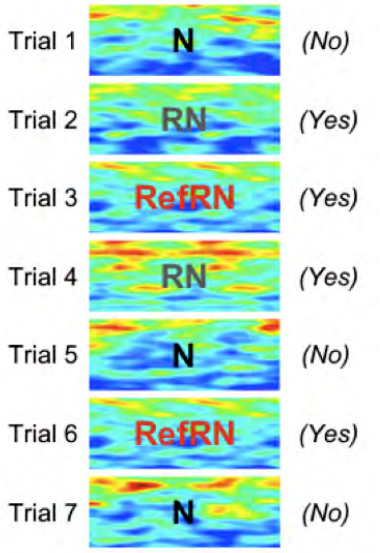
While the results for medial temporal lobe structures such as the hippocampus do not appear compatible with "grandmother cell" coding, it could be that the hippocampus and cortex differ and that the hippocampus may not be involved in very long-term memories at all. Indeed, I recently made the alternative proposition that hippocampal neurons may be keeping track of stimuli that have been seen in the relatively recent past – possibly over a period of a few weeks or months (Thorpe 2011). Thus a neuron in the hippocampus that responds selectively to "Jennifer Anniston" today, might respond to something completely different in a couple of years time. This leaves open the possibility that the situation in neocortex could be radically different, especially if it is the neocortex that is involved in storing the sorts of extremely long-term memories that interest us here. Perhaps the hit rate for finding highly selective cells would be very much lower than in hippocampus – and that many neocortical cells could be virtually impossible to activate. The fact is that we could not possibly know, because by definition, a neuron that never fires cannot be studied using conventional techniques.

Robust memory formation in humans

We have already seen that STDP-based learning could potentially provide a powerful way of creating selectivity to stimuli that occur repeatedly. But what evidence is there that this sort of automatic learning of repeated stimuli occurs in humans? Interestingly, some recent work that I did with Trevor Agus and Daniel Pressnitzer has provided direct evidence for this sort of learning (Agus et al 2010). In this study, we presented participants with meaningless auditory noise patterns 1 second in duration. For half the stimuli, the first and second half seconds of the stimulus were identical, and the task of the participants was to report whether the two halves matched or not (see Figure 1 for more detail). It is a difficult task to perform and with new stimuli, participants perform only marginally above chance. However, some of the target stimuli occurred repeatedly during the testing session. For these repeated stimuli, performance rapidly improved over the first 10 to 20 trials demonstrating that the participants were learning about the stimuli, even though they were unaware that any of the targets were going to reoccur. Intriguingly, not all the targets could be learned, but for those that were, performance reached almost 100% correct (see Figure 2).

One of the most remarkable features of the results was that when participants returned to the lab to be tested in a second session 2-3 weeks after the original training, performance was almost as good from the very first trials of the new session. Clearly, the subjects had formed robust memories for the stimuli that lasted 2-3 weeks at least. In the present context it is very important to realise that because of the nature of the stimuli (meaningless auditory noise patterns) it would be effectively impossible for the participants to rehearse the stimuli in the intermediate period. Thus, the results provide one of the very few situations where it can be demonstrated that a sensory stimulus can be memorised over an extended period **without the possibility of rehearsal**.

Figure 1. Samples of the Gaussian noise stimuli used in the Agus et al (2010) study, illustrated here as schematic spectrograms containing random amplitude fluctuations across time and frequency. Listeners were asked to detect those trials that contained a repetition. The noise (N) trials were formed from segments of noise, so the correct response would be ''No'' repetition. The repeated-noise (RN) trials were formed from the seamless repetition of a half-duration segment of noise, for which the correct response would be ''Yes.'' The N and RN trials were generated afresh for each trial. The reference repeated-noise (RefRN) trials also contained a repetition but, importantly, the exact same reference noise sample was used over several trials. The fact that performance improves dramatically within a few tens of presentations is strong evidence that the subjects can form long-term memories for these meaningless stimuli.



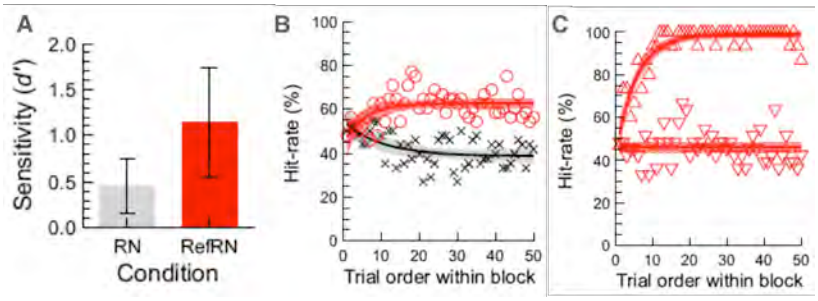


Figure 2. A. Difference in performance for detecting novel targets (RN) and targets that repeated during the session (RefRN). **B.** Variation in performance during the first testing session for novel targets (in grey) and re-occurring targets (in red). **C.** The data in B split to show performance for stimuli that could be learned (upward arrows) and those where learning failed (downward arrows)

Computational modeling

Modelling the learning of meaningless auditory noise

What might be happening at the neural level during this sort of psychophysical learning task? Our recent neural modelling studies have provided a possible explanation that requires nothing more sophisticated than one or more leaky integrate and fire neurons equipped with an STDP learning rule.

Most of our simulation work has used the standard STDP rule in which only synaptic inputs that fire within a short interval of a post-synaptic spike are modified – those that fire before the spike are potentiated, those that fire afterwards are depressed (see the red and blue solid lines in Figure 3). However, with my student Olivier Bichler, we have recently discovered that an even simpler formulation has some remarkably interesting properties. In our new implementation, also illustrated in Figure 3, the effect of a postsynaptic spike is to decrease the synaptic weights of all inputs, whether they have fired or not. The only exception are any inputs that fired just before the postsynaptic spike – the hatched red region in Figure 3. This very simple rule turns out to be very powerful. By appropriately adjusting the ratio between potentiation and depression, we have found that STDP-based learning provides a remarkably powerful way to generate selective neuronal responses.

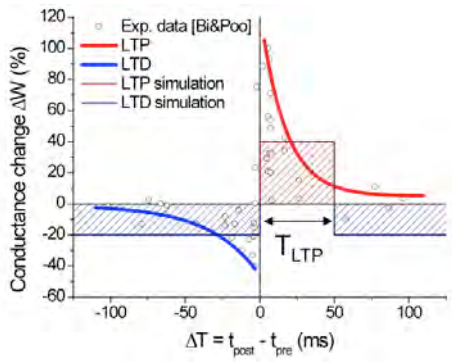


Figure 3. Comparison of different types of STDP learning rules. In the conventional STDP rules (based on Bi and Poo's experimental results – circles), only synapses activated within a certain time difference of the postsynaptic spike are modified – those firing before the spike are potentiated (solid red line), whereas those firing afterwards are depressed (solid blue line). In our new proposal, **ALL** synapses are depressed when there is a postsynaptic spike (blue hatched region) **EXCEPT** those that fired just before the postsynaptic spike (red hatched region). Note that there is effectively no "temporal window" for the depression since it affects all synaptic inputs, whether they fire or not.

Very recently, we used this modified STDP rule to model the learning of the same auditory noise stimuli used for the psychophysical studies (Bichler et al, in preparation). We used a simple model of the auditory periphery to generate spiking activity in 3000 fibres, each roughly corresponding to a frequency-tuned fibre in the auditory nerve. Driving the model with Gaussian noise will therefore generate a barrage of spikes in this artificial auditory nerve. The outputs of these 3000 afferents were

then fed to a single leaky integrate and fire neuron equipped with our modified STDP learning rule. The results were spectacular.

As illustrated in Figure 4, we initially presented 400 seconds of continuous auditory noise and recorded the response of the neuron. **Panel A** shows that the neuron maintained a continuous irregular firing rate of around 2 spikes per second throughout this period with no sign of modification. From then on, we started introducing 1 second blocks of noise at irregular intervals that corresponded to the RefRN stimuli used in the psychophysical experiments, with 2 identical 500 ms noise stimuli back to back. As you can see from **Panel B** which shows the very first times these training stimuli were shown, the neuron fires twice during each training stimulus, but it still fires to the background noise. However, within a few minutes, and after only 25 presentations of the training pattern, the neuron is now firing 4 spikes during each pattern, but there are no longer spikes to the background noise (**Panel C**). In other words, the neuron has become highly selective to the repeating pattern. At the end of this training period, we then presented another 400 seconds of continuously renewed noise. As illustrated in **Panel D**, there were no more spikes at all from the neuron. To all extents and purposes the neuron is now a "grandmother cell" because although it will respond very selectively when its preferred stimulus is presented again, it will remain totally silent when presented with noise.

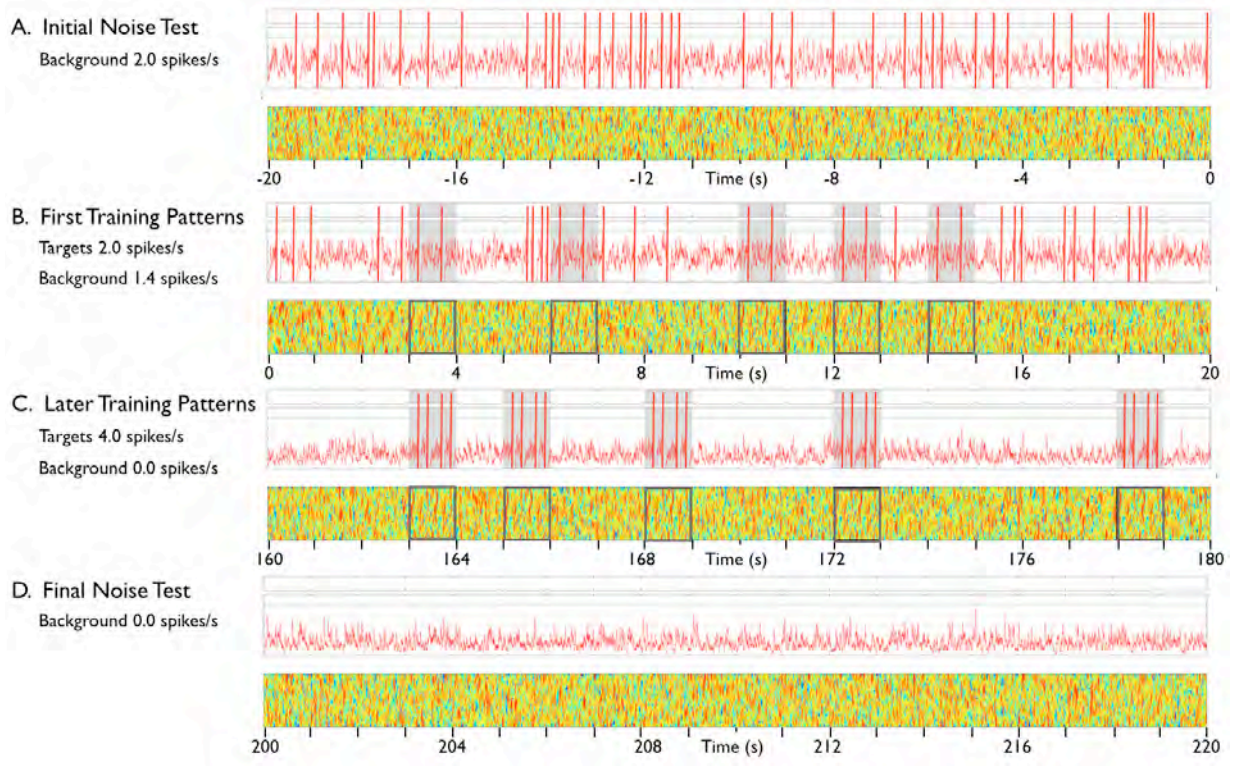


Figure 4. Generation of a "grandmother cell" selective to a repeating auditory noise pattern using a simple STDP rule. In each panel, there are two traces. The upper trace (red line) shows the membrane potential of a single neuron receiving inputs from 3000 simulated auditory nerve fibres. The lower multicoloured image shows a spectrogram of the frequency content of the auditory noise stimulus. The training patterns are marked by the grey rectangles in panels B and C. See text for more details.

The implications of these modelling results are profound. Note that the neuron has become extremely selective, but there is nothing "special" about the particular stimulus to which the neuron has learned to respond. Indeed, it would learn to respond to any arbitrarily selected pattern, as we had already noted in previous studies (Masquelier et al 2008). The only thing needed to force the neuron to become selective was to present the input pattern repeatedly. Furthermore, we only needed around 25 presentations to make the neuron highly selective. Could it be that this rule of thumb, that 25 repetitions are enough to store a sensory pattern in neuronal permastore, is a general rule of sensory learning? It is this key idea that we will be testing explicitly in the current project.

In fact, there are a whole range of predictions that can be made using this simple modelling approach. For example, the model would predict that the probability that a neuron becomes selective to a repeating pattern is not simply a question of how many times the stimulus is repeated. We predict that the delay between the repeats will also be critical. Specifically, learning will be optimal if the repeats of the stimulus are grouped closely together. In contrast, if each repetition was separated by

several tens of seconds, and the gaps filled with continuous noise, the neuron would have difficulty learning the stimulus because effectively the spikes generated to the background noise will cause the neuron to “unlearn” the pattern of synaptic connections that corresponds to the stimulus. We can thus use the model to derive specific predictions about how different training protocols should influence the ability of the system to learn - predictions that can be directly tested in humans using psychophysical methods. And, as mentioned earlier, such findings could have important implications for the development of optimal teaching methods.

Unsupervised learning of dynamic visual inputs

There is nothing special about the auditory noise stimuli. In other studies my student Olivier Bichler has been using the exact same approach to train neurons to respond selectively to visual stimuli (Bichler et al 2011, 2012). Rather than using the output of a simulated auditory nerve, we used data obtained with an asynchronously firing retina chip developed by Tobi Delbruck and colleagues in Zurich (Delbruck et al 2010; Lichtsteiner et al 2008). The chip has 128*128 pixels, and for each pixel, there are effectively two "neurons" – one that fires a spike when the local luminance increases (similar to the "ON" channels in a biological retina), and the other than fires when the luminance decreases (corresponding to an "OFF" response). There are various datasets available for the chip, including one corresponding to a few minutes of traffic on a six-lane freeway in California. Bichler took the spikes generated by the retina chip, and fed them into two layers of STDP-equipped neurons, as illustrated in figure 5. As in the auditory noise learning example, the neurons learn to respond to patterns of spikes that occur repeatedly in the input. In this case, this meant that the neurons in the first layer learnt to respond to cars moving at specific locations in the image. Even more remarkably, the second layer neurons spontaneously learned to "count" the cars going past on each lane of the freeway. The truly amazing feature of this learning is that is completely unsupervised – no-one had to tell the system what it was supposed to do. The system spontaneously learned to count cars on different lanes because these are the "events" in the inputs that repeat reliably.

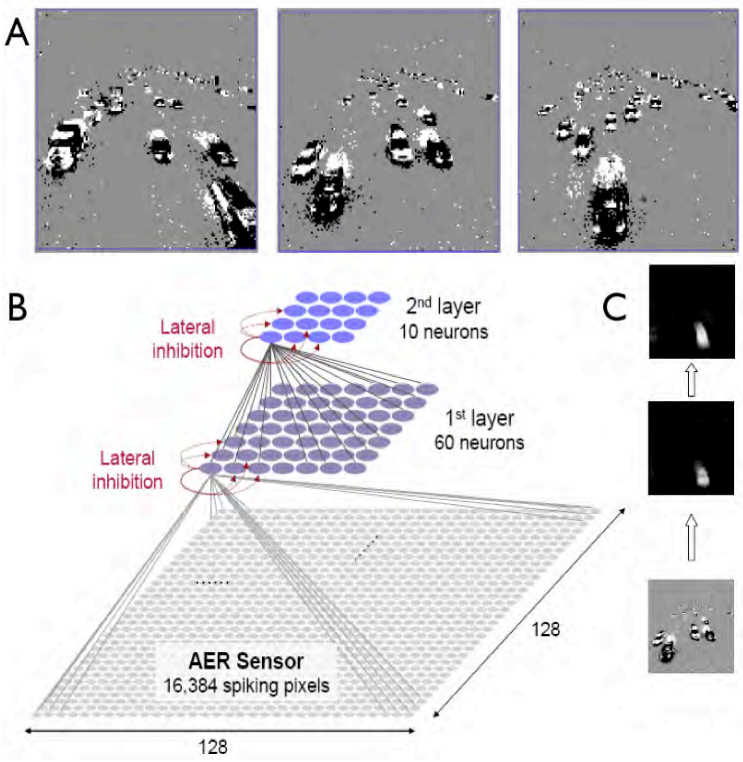


Figure 5: Unsupervised learning in a two layer network connected to an asynchronously firing spiking retina chip. **A.** Illustration of "images" corresponding to the spikes from "ON" cells (white dots) and "OFF" cells (black dots) in response to cars travelling on a freeway. **B.** Architecture of the network. The AER Sensor has 128*128 pixels each with two polarities. These feed to 60 first layer neurons that are in turn connected to 10 second layer neurons. **C.** Neurons in the first layer develop receptive fields that are tuned to cars moving at specific locations, whereas some of the second layer neurons spontaneously learn to "count" cars travelling on specific lanes (Bichler et al 2011, 2012).

This example makes another point. While the auditory noise learning example mentioned earlier was extremely simple - there is just a single neuron receiving spiking inputs directly from the auditory nerve - nothing prevents us from developing much more sophisticated architectures with large numbers of neurons arranged in multilayer hierarchical structures. With a simple spiking retina projecting to just 60 neurons in a first layer, which in turn project to 10 neurons in the second layer, we already have an architecture that can spontaneously learn to count cars. Imagine the potential of a multilayer hierarchy for auditory processing in which the first layer neurons learn to code the low-level acoustic features of

the input! What might we find in subsequent layers? Perhaps at one layer we would find that neurons become selective to phonemes, while those even higher up in the system might become selective to words or even whole phrases.

Clearly, this modelling approach has enormous potential and could be applied to a whole range of problems. Indeed, one of the aims of the current project will be develop models based on networks of spiking neurons for a whole range of sensory learning situations, including not just auditory and visual stimulation, but also multimodal processing. In every case, we plan to test the ideas and predictions with experimental studies in human subjects, experiments that can be combined with a range of brain imaging techniques including fMRI and EEG recording in order to pin down the brain structures where the changes are occurring.

Towards bio-inspired hardware

The M4 project will not only attempt to link experimental data on the formation of long-term sensory memories with detailed spike-based neural network models. We are also convinced that the ideas being developed here can potentially be implemented in electronic hardware. In collaboration with Christian Gamrat and other colleagues from the CEA in Saclay and Grenoble we are looking into the idea that STDP could be implemented using components called "memristors". These are semiconductor devices that have a resistance that can be changed by applying a voltage that exceeds a certain threshold. Because of this, they can be programmed in a way that is exactly analogous to the way that STDP can program the weights of biological synapse. The principle is illustrated in Figure 6.

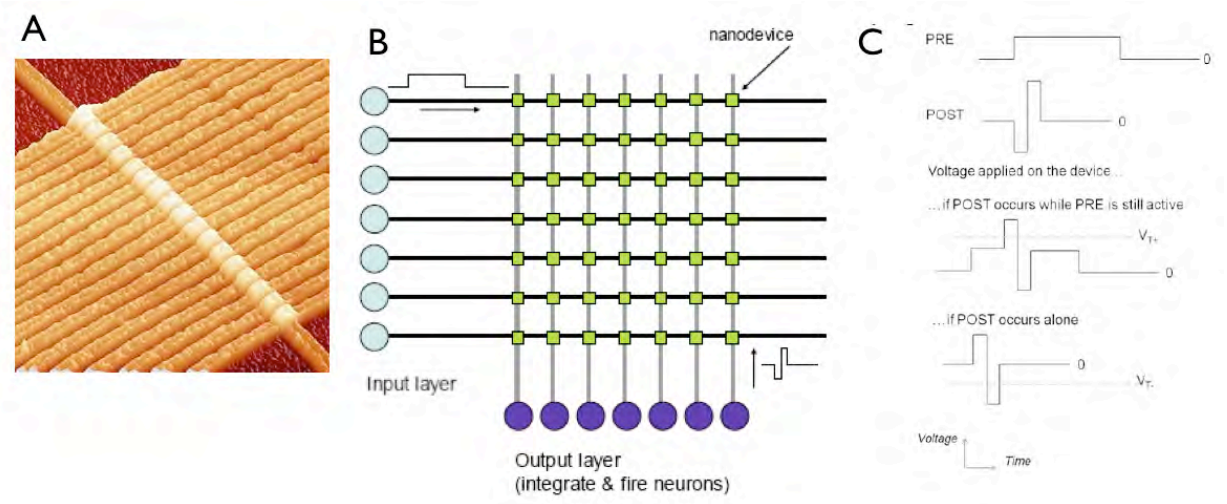


Figure 6: Implementing STDP based learning in hardware. **A.** Highly magnified view of a set of memristor connections (Strukov et al 2008) **B.** Implementing a set of "synapses" between neurons in an input layer and neurons in an output layer in which the resistance of the connections can be programmed by choosing the input and output pulse shapes appropriately. **C.** Illustration of how a suitable choice of pre- and post-synaptic pulse shapes can mean that when a pre-synaptic pulse precedes the post-synaptic pulse, the combined voltages exceed a threshold voltage and cause a reduction in resistance.

The potential for this sort of memristor based learning architecture is enormous. In principle it should be possible to build electronic synapses that are a mere 10 nanometers in diameter, opening up the possibility of building chips wit as many as 100 million synapses per square centimetre of silicon. Furthermore, with 3D-stacking technologies, it may be possible to build a system with the same number of programmable synapses as the human brain. Such a system would be a radical departure from conventional approaches that attempt to simulate the brain using conventional computer hardware. In this case, there would be nothing equivalent to the CPU of a conventional computer. Instead, the system would function very much as the human brain does – but constantly reprogramming connections and forming neurons that become selective to significant stimuli. Indeed, like the human brain, such a system could learn about a stimulus at one time, and still be able to recognise it decades later.

Tracking the development of selectivity at the single unit level in humans

With two researchers working in my group, we are currently setting up to perform single unit recording in human epileptic patients undergoing presurgical examination. Emmanuel Barbeau already has considerable experience in intracerebral recordings of local field potentials (Barbeau et al 2008, Kirchner et al 2009, Maillard et al 2011) and Leila Reddy did her doctoral thesis in Itzak Fried and Christof Koch's group in California responsible for the recordings of highly selective and invariant neurons in the human medial temporal lobe - including the famous "Jennifer Anniston" cell (Quian Quiroga et al 2005, Reddy et al 2006). In the experiments that we propose to perform in the current project we aim to use single unit recording to test one of the major predictions of our modelling studies.

Of course, ideally it would be good to be able to prove the "grand-mother cell" hypothesis by looking at the responses of individual cortical neurons and testing their selectivity. However, as we have argued, it is possible that there are substantial numbers of neurons in cortical structures that simply never fire spikes (Shoham et al 2006). This makes a conventional approach of going into a cortical structure and "fishing" for grandmother cells highly problematic, and such a strategy might well be doomed to failure.

Our approach will be different. It is well known that when a visual stimulus is repeated, the neuronal response to the second presentation is typically reduced – a phenomenon known as "repetition suppression" (Desimone, 1996). Such effects have often been reported in monkeys (e.g. De Baene & Vogels, 2010), and have also been seen for single neurons in the human medial temporal lobe (Pedreira, et al., 2010). Various mechanisms have been proposed to explain the phenomenon, including sensory fatigue, selective sharpening, faster and briefer processing of the repeated stimuli, (Grill-Spector, Henson, & Martin, 2006).

However, we have an alternative suggestion based on our neural network simulation work. Our proposition is that while most neurons respond less strongly to repeated presentation of the same stimulus, there may be a few neurons that are actually becoming more selective to the repeating stimulus, and that these cells produce inhibition that reduces the global level of activation. As a consequence, the average neuron will respond less, and the overall fRMI or ERP signal is typically reduced. The idea is illustrated in panels A and B of figure 7. This pattern is a direct prediction of our modelling studies which propose that only one cell needs to learn any given stimulus, and its inhibitory connections will reduce the responses of other local neurons.

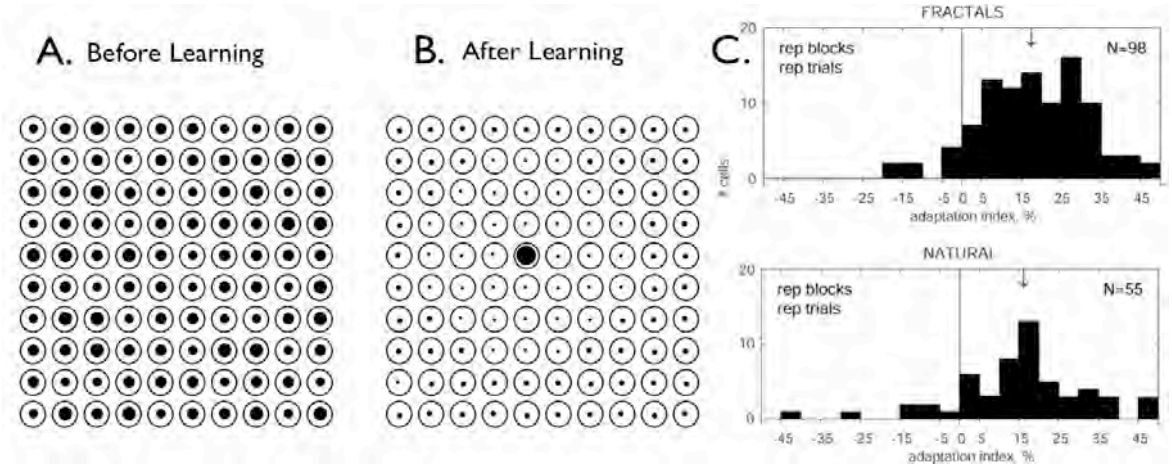


Figure 7. Panels A and B show how a hypothetical population of 100 cells might respond before and after learning of a new stimulus. The size of the central disk is supposed to represent the size of the neuronal response. Initially, most of the neurons may be activated relatively weakly (A), but after learning, one particular neuron has become selective and responds strongly. Because of intracortical inhibitory circuits, this could reduce the level of responding in neighbouring neurons (B). Panel C shows the distribution of an adaptation index for neurons in monkey inferotemporal cortex between the first and second presentation of a previously unseen fractal pattern (upper panel) or natural image (lower panel - the data are redrawn from a recent study (Kaliukhovich & Vogels, 2011).

Interestingly, while nearly all studies only describe repetition suppression, a recent study revealed that there are a few cells that appear to increase their responses (Kaliukhovich & Vogels, 2011). Panel C of Figure 7 shows that out of a sample of around 150 neurons recorded from monkey inferotemporal

cortex, there were a handful of neurons that responded more strongly to the second presentation, with one cell responding roughly 45% more. This effect is extremely interesting, because the handful of neurons that show increased responses after repeated exposure could be precisely those involved in learning to recognise the stimuli. Unfortunately, the study only looked at the responses to the first and second presentations, so we have no way of knowing what would have happened if the neuron had been shown its preferred stimulus several times. Would the response have become even stronger?

We will therefore investigate this question by recording the responses of single cortical and subcortical neurons to repeated presentations of previously unseen images. In the first session, we will test a set of neurons with a large number of images (typically between 100 and 500 in a given session). Each image will be shown at least twice, and we will expect that the vast majority of cells will respond less to the second and subsequent presentations. However, we will be specifically looking for any cells that showed an increase in response between the first and second presentations. If we find one, the second testing session (later on during the day) will intermingle additional presentations of that particular stimulus with a range of other images. Based on our neural network simulation studies (such as the one illustrated in Figure 4), we would predict that if a neuron increased its response between the 1st and 2nd presentations, it should become even more responsive with repeated presentations. Furthermore, its spontaneous activity should get lower and lower. If this result holds, then we could potentially watch while a neuron becomes a “grandmother cell” - becoming more an more selective to a specific visual stimulus while losing its spontaneous activity. As a consequence, that cell could potentially become sufficient inactive to allow it to become the instantiation of the memory for that particular stimulus.

Other Objectives

The current proposal will build on our previous work on the formation of robust memories for auditory noise and our modelling results but we will also be developing some radically new research programs that will make use of web-based testing methods. We will also be using a range of methods for analysing the underlying brain structures that will include fMRI, EEG and intracerebral recording studies. The overall work-plan contains several components.

Studies on long-term memories for meaningless auditory noise stimuli. This will extend our previous work by answering the following questions. 1) Can the memory traces survive for longer than 2-3 weeks, and what is the form of the forgetting function relating performance with the interval since learning. 2) What is the function linking memory strength with the number of presentations during the training session? 3) Can we find evidence for learning related changes in the brain using imaging methods such as fMRI and auditory Event-Related Potentials.

Studies on long-term memories for meaningless visual noise stimuli. Given the surprising success of the auditory noise learning paradigm, we will perform equivalent experiments to see if similar results can be obtained in the visual modality too. Again, these experiments can be combined with imaging.

Studies on verbal labelling of auditory, visual and multimodal material. We will also perform other lab based experiments that will extend our studies of memories for meaningless auditory and visual noise to more realistic conditions. Specifically, we will investigate the ability of subjects to associate an arbitrary verbal label with short auditory, visual or visual and auditory stimuli. In particular, we will derive the function relating the number of training experiences with the strength of the memory trace and how well it survives long delays extending up to 2-3 years.

Studies on the recognition of TV and Radio themes from the 50s and 60s. Here again, our basic aim is to get hard information about how the probability of being able to remember a particular radio or TV theme varies as a function of the number of times it was experienced decades ago. Unlike other old material, we can obtain information about precisely how many times a particular program was broadcast. This, together with participants own reporting should allow us to piece together a coherent story. One particularly straightforward outcome would be if all the different acquisition functions – those obtained with meaningless (and unrehearsable) noise stimuli, those obtained with artificial "clips" in the laboratory, and the real-world data obtained with the TV and Radio themes – all had the same basic shape. If that was the case, it would provide very strong support for the view that the underlying mechanisms may be the same for all these sensory memories.

b. Methodology

In this section I will just give details of some of the more specific experimental procedures that we will be using in the current project.

Memories for meaningless auditory and visual noise stimuli

The experiments on the learning of meaningless auditory noise will use the same sorts of stimuli that have already been used for the psychophysical experiments (Agus et al 2010) and for the modelling study mentioned earlier. But we also plan to perform analogous experiments in the visual domain by presenting a sequence of rapidly presented random dot patterns lasting one second (see Figure 8). Again, the task for the subject is to decide whether the first and second halves of the sequence are identical. As with the auditory noise learning experiment, we expect that performance with a repeated-noise (RN) trials will be very poor. However, if learning in the visual system works in a similar way to the auditory system, we may well find that performance improves for stimuli that reoccur during the training session.

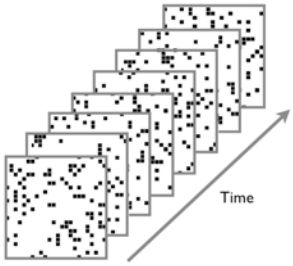


Figure 8. Can subjects form sensory memories for meaningless visual noise? Subjects will be shown sequences of random dot patterns typically lasting 1 second. On half the trails, the first and second halves are the same, and the task is to respond "Yes" or "No" depending on whether there is a repeat or not. Some of the patterns reoccur, and the memorisation (if it occurs) is visible as an increase in sensitivity with repeated presentation

Brain activation associated with memory formation

We already have some pilot data showing that various brain structures show differential activation following training in the auditory noise learning task. Figure 9 shows preliminary data from 8 subjects showing that at least three brain areas showed higher responses to stimuli that were successfully learned (left hippocampus, bilateral thalamus and auditory cortex). We also have pilot data (not illustrated) indicating an increased auditory ERP response at a latency of 200 ms following stimulus onset. Such results demonstrate that we will be able to study which brain structures are associated with the learning.

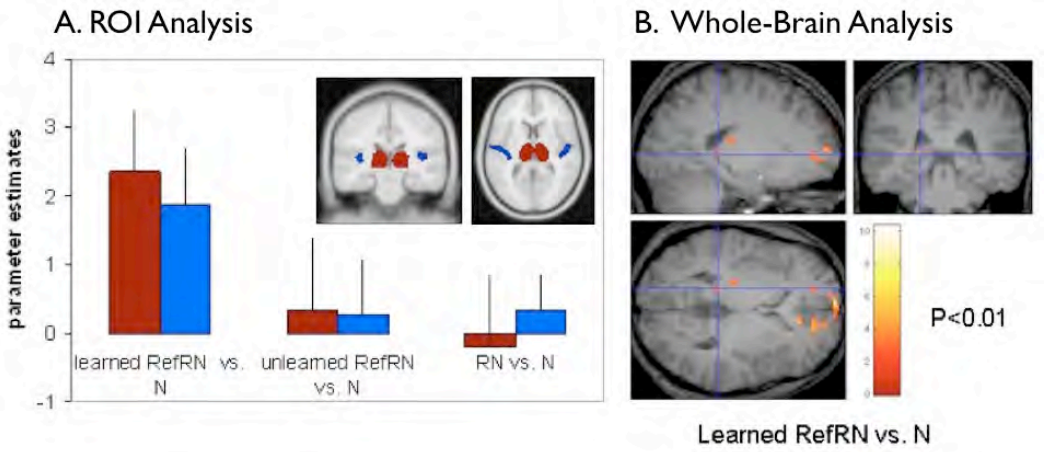


Figure 9. Pilot fMRI from 8 subjects. Subjects underwent a pre-training session where they were asked to categorize 1-s Gaussian sounds as repeated (RN) or not (N). Following analysis of the pre-training data for each subject, RefRNs were classified as 'learned RefRN' when subject's performance was above 90%, and as 'unlearned RefRN' when performance was under 60% (based on average categorization performance on newly heard RN sounds). For fMRI experiments, each trial included 9 s without image acquisition followed by 3-s acquisition. During the 9-s (silent) period, the sound was presented at the 6th second through headphones. Group data were examined using whole-brain analysis, and showed higher activity in the left posterior hippocampus for previously learned RefRNs (right panel). ROI analysis was conducted in bilateral thalamus (red ROI) and primary auditory (blue ROI) regions and showed higher activity for learned RefRN sounds (left panel).

Learning to associate verbal labels with new visual and auditory stimuli

These experiments will use a different approach in which we require subjects to learn an arbitrary verbal label for each stimulus during the initial training session. The labels will be a set of meaningless pseudo-words (for example "Winolit", "Bajat", "Ramikon" , etc.) that can be generated with publicly available software. The basic structure of a training trial is illustrated in figure 10.

After an inter-trial interval, one of the randomly selected labels is presented at the centre of the screen for 1 second. Then the training stimulus is presented for a duration that can vary between roughly 500 ms and 3 seconds (we will choose the precise values following some pilot experiments). The training stimuli can be of various types. It could be (a) a purely auditory stimulus, such as a segment of music, (b) a static image, (c) a video sequence or (d) a combination of audio and video. The label then reappears for a further second and the subject's task is to try to remember the label that goes with each training stimulus.

As with the experiments performed with meaningless auditory and visual noise, we will be able to combine these behavioural experiments with ERP and fMRI imaging. Again, such data may well provide key insights into the brain mechanisms involved in this sort of learning. Specifically, we will be able to look for brain activation patterns that scale with the number of repetitions during the training session. If we were able to find a brain structure where the size of the fMRI activation varied in the same way as the subjects performance in the memory task, this would provide a powerful way of localising areas that could be involved in memory formation.

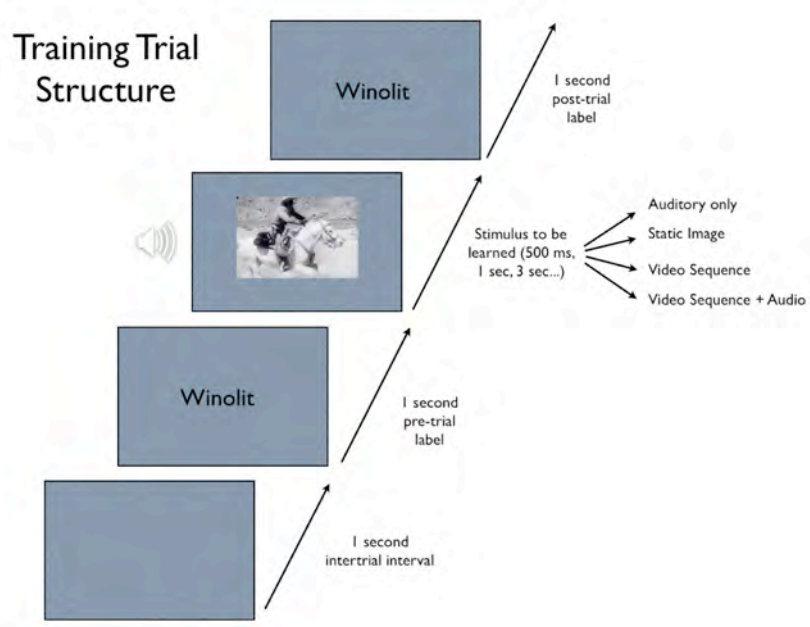


Figure 10. Structure of a training trial. After a short inter-trial interval, participants are shown a verbal label for 1 second. This is then followed by a clip which can be of one of four types –Audio alone, Static image, Video alone, or Video + Audio. The label is then shown again for a second before starting another trial.

Memories for very remote visual and auditory stimuli

Subjects will be asked to fill in a questionnaire that provides details about their age, sex, nationality, whether or not they had a television at home when they were young, and roughly how much time they spent watching the television. They will then be shown a brief clip lasting a few seconds from an old television or radio program and asked if they can name it. Their verbal responses are recorded and analysed offline (see Figure 11 for an example).

Figure 11. Screen shot from a prototype of the web-based application that we intend to use to test the ability of subjects to identify old television and radio sequences. For more examples, see the site at **http://cerco.ups-tlse.fr/elma/**. This particular example is a clip from the intro to a program called "Space Patrol" that had 39 episodes that were broadcast between 7th April 1963 and 11th June 1964. It has never been rebroadcast on the television. So far, no-one except people who saw the program at the time have been able to identify it, demonstrating that the material does not form part of widely available cultural information.

Our aim is to test large numbers of people with similar material, using a web-based testing procedure. They would then be shown a clip from the opening sequence of a television program originally broadcast in the 1950s, 1960s or 1970s. If the viewer is able to recognise the program and provide some explicit information about it (for example, "Lone Ranger", "Bewitched", "Bonanza"), then the result is considered a hit. In that case the participant would be asked a series of questions such as (i) was the program one that they used to watch? (ii) have they seen the program in the intervening period, and if so, how many times and when? The interesting cases will be those participants who were able to recognise and name the program, but had not seen the program since. In those cases, it is important to try to estimate the number of times that the person saw the program. Thanks to information available from institutions like the INA in France and the BBC in the UK, it should be possible to provide the precise number of times that a particular program was broadcast (39, in the case of "Space Patrol"). It seems reasonable to assume that people who were watching television in 1963–4 might have a reasonably reliable idea about roughly how many times they saw it. This value will range from zero ("I've never seen it"), to 1 ("I remember seeing it once but didn't like it"), to around 20 ("I think I may have only seen the second series"), to close to 39 ("It was my favourite TV program, and I think that I never missed an episode").

As stated earlier, the key result that we hope to get from this study is a graph that plots the probability that a particular program can be recognised on the basis of a brief clip as a function of the number of times that the program was seen. Obviously, it is likely that there will be considerable individual variation between participants in their ability to remember such old material, but our belief is that with a sufficiently large number of participants, the results will be reliable. Importantly, the use of large numbers of participants from different backgrounds, different locations and with different ages allows these experiments to have a number of built-in controls that increase the reliability of the data. For example, consider a program broadcast in 1963-4 that was reliably recognised by a substantial percentage of people who were between 6 and 15 years old in 1963. This would in principle suggest that they had been able to retain the memory for 47 years. However, if there were cases of people who were capable of recognising the program but who were not even born in 1963 (aged 45 or less), this would make it highly likely that the program had been rebroadcast at some point, and would mean that the data point should probably be discarded.

Final Comments

The M4 proposal starts from an idea that most adults would accept – the idea that we can recognise things that we have not seen for decades, even without having the opportunity to reactivate the memory. A first challenge is to prove that this phenomenon really does exist, and to do this we will use a highly innovative research program using web-based methods that could potentially allow us to test very old memories in tens of thousands of people. A second challenge is to analyse the conditions required forthe formation of these extremely long-term memories - research that could have major implications for education. The experimental work will be coupled with our neural network simulations that have already proved very effective at modelling learning for auditory noise stimuli. Together, these different sets of data should allow us to fully test a revolutionary idea about how long term memories are stored – an idea that proposes both the existence of "grandmother cells" and neocortical "dark matter". Finally, we will use these ideas to develop new computational paradigms that can be implemented in revolutionary hardware that can mimic the way the brain stores long-term memories.

Cited References

Agus TR, Thorpe SJ, Pressnitzer D. 2010. Rapid Formation of Robust Auditory Memories: Insights from Noise. *Neuron* 66:610-8

Arshavsky YI. 2006. "The seven sins" of the Hebbian synapse: Can the hypothesis of synaptic plasticity explain long-term memory consolidation? *Prog Neurobiol* 80:99-113

Azevedo FA, Carvalho LR, Grinberg LT, Farfel JM, Ferretti RE, et al. 2009. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J Comp Neurol* 513:532-41

Bahrick HP. 1983. The cognitive map of a city - 50 years of learning and memory. *Psychology of Learning and Motivation-Advances in Research and Theory* 17:125-63

Bahrick HP. 1984b. Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *J Exp Psychol Gen* 113:1-29

Bahrick HP, Bahrick PO, Wittlinger RP. 1975. Fifty years of memory for names and faces: A cross-sectional approach. *J Exp Psychol Gen* 104:54-75

Barbeau EJ, Taylor MJ, Regis J, Marquis P, Chauvel P, Liegeois-Chauvel C. 2008. Spatio temporal Dynamics of Face Recognition. *Cereb Cortex* 18:997-1009

Barlow HB. 1972. Single units and sensation: a neuron doctrine for perceptual psychology? *Perception* 1:371-94

Baum EB, Moody J, Wilczek F. 1988. Internal Representations for Associative Memory. *Biol Cybern* 59:217-88

Bayley PJ, Gold JJ, Hopkins RO, Squire LR. 2005. The neuroanatomy of remote memory. *Neuron* 46:799-810
Bayley PJ, Hopkins RO, Squire LR. 2006. The fate of old memories after medial temporal lobe damage. *J Neurosci* 26:13311-7
Bi GQ, Poo MM. 2001. Synaptic modification by correlated activity : Hebb's postulate revisited. *Ann Rev Neurosci* 24:139-66
Bichler O, Querlioz D, Thorpe SJ, Bourgoin JP, Gamrat C. 2011. Unsupervised Features Extraction from Asynchronous Silicon Retina through Spike-Timing-Dependent Plasticity. In *2011 International Joint Conference on Neural Networks*, pp. 859-66
Bichler O, Querlioz D, Thorpe SJ, Bourgoin JP, Gamrat C. 2012. Extraction of temporally correlated features from dynamic vision sensors with spike-timing-dependent plasticity. *Neural Networks* Epub ahead of print
Blanche TJ, Spacek MA, Hetke JF, Swindale NV. 2005. Polytrodes: high-density silicon electrode arrays for large-scale multiunit recording. *J Neurophysiol* 93:2987-3000
Bowers JS. 2009. On the biological plausibility of grandmother cells: Implications for neural network theories in psychology and neuroscience. *Psychol Rev* 116:220-51
Bowers JS. 2010. More on grandmother cells and the biological implausibility of PDP models of cognition: a reply to Plaut and McClelland (2010) and Quian Quiroga and Kreiman (2010). *Psychol Rev* 117:300-8
Bowers JS. 2011. What is a grandmother cell? And how would you know if you found one? *Connection Science* 23:91-5
Brady TF, Konkle T, Alvarez GA, Oliva A. 2008. Visual long-term memory has a massive storage capacity for object details. *Proc Natl Acad Sci U S A* 105:14325-9
Bredt DS, Nicoll RA. 2003. AMPA receptor trafficking at excitatory synapses. *Neuron* 40:361-79
Bruck M, Cavanagh P, Ceci SJ. 1991. Fortysomething - Recognising faces at one 25th reunion. *Mem Cogn* 19:221-8
Buckner RL, Wheeler ME. 2001. The cognitive neuroscience of remembering. *Nat Rev Neurosci* 2:624-34.
Caporale N, Dan Y. 2008. Spike timing-dependent plasticity: a hebbian learning rule. *Annu Rev Neurosci* 31:25-46
Conway MA, Bekerian DA. 1987. Organization in Autobiographical Memory. *Memory & Cognition* 15:119-32
Conway MA, Pleydell-Pearce CW, Whitecross S, Sharpe H. 2002. Brain imaging autobiographical memory. In *Psychology of Learning and Motivation: Advances in Research and Theory*, ed. BH Ross, pp. 229-63
Conway ARA, Skitka LJ, Hemmerich JA, Kershaw TC. 2009. Flashbulb Memory for 11 September 2001. *Applied Cognitive Psychology* 23:605-23
De Baene W, Vogels R. 2010. Effects of adaptation on the stimulus selectivity of macaque inferior temporal spiking activity and local field potentials. *Cereb Cortex* 20:2145-65
Delbruck T, Linares-Barranco B, Culurciello E, Posch C, Ieee. 2010. Activity-Driven, Event-Based Vision Sensors. In *2010 Ieee International Symposium on Circuits and Systems*, pp. 2426-9
Desimone R. 1996. Neural mechanisms for visual memory and their role in attention. *Proc. Natl. Acad. Sci. U. S. A.* 93:13494-9
Dudai Y. 1997. How big is human memory, or on being just useful enough. *Learn Mem* 3:341-65
Ebbinghaus H. 1913. *Memory: A contribution to experimental psychology*: Teachers college, Columbia university
Feldman DE. 2009. Synaptic mechanisms for plasticity in neocortex. *Annu Rev Neurosci* 32:33-55Grill-Spector K, Henson R, Martin A. 2006. Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci* 10:14-23
Finelli LA, Haney S, Bazhenov M, Stopfer M, Sejnowski TJ. 2008. Synaptic learning rules and sparse coding in a model sensory system. *PLoS Comput Biol* 4:e1000062
Gross CG. 2002. Genealogy of the "Grandmother Cell". *Neuroscientist* 8:84-90
Guyonneau R, Vanrullen R, Thorpe SJ. 2005. Neurons Tune to the Earliest Spikes Through STDP. *Neural Comput* 17:859-79
Hahnloser RH, Kozhevnikov AA, Fee MS. 2002. An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* 419:65-70.
Haist F, Gore JB, Mao H. 2001. Consolidation of human memory over decades revealed by functional magnetic resonance imaging. *Nat Neurosci* 4:1139-45
Hebb DO. 1949. *The Organization of Behaviour : A Neuropsychological Perspective*. New York: Wiley
Hirst W, Phelps EA, Buckner RL, Budson AE, Cuc A, et al. 2009. Long-Term Memory for the Terrorist Attack of September 11: Flashbulb Memories, Event Memories, and the Factors That Influence Their Retention. *Journal of Experimental Psychology-General* 138:161-76
Hopfield JJ. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci U S A* 79:2554-8.
Jensen O, Bonnefond M, Vanrullen R. 2012. An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends Cogn Sci* 16:200-6
Jortner RA, Farivar SS, Laurent G. 2007. A simple connectivity scheme for sparse coding in an olfactory system. *Journal of Neuroscience* 27:1659-69
Katona G, Szalay G, Maak P, Kaszas A, Veress M, et al. 2012. Fast two-photon in vivo imaging with three-dimensional random-access scanning in large tissue volumes. *Nature Methods* 9:201-8
Kirchner H, Barbeau EJ, Thorpe SJ, Regis J, Liegeois-Chauvel C. 2009. Ultra-rapid sensory responses in the human frontal eye field region. *J Neurosci* 29:7599-606
Klimesch W, Fellinger R, Freunberger R. 2011. Alpha oscillations and early stages of visual encoding. *Front Psychol* 2:118
Konkle T, Brady TF, Alvarez GA, Oliva A. 2010a. Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *J Exp Psychol Gen* 139:558-78
Konkle T, Brady TF, Alvarez GA, Oliva A. 2010b. Scene memory is more detailed than you think: the role of categories in visual long-term memory. *Psychol Sci* 21:1551-6
Kaliukhovich DA, Vogels R. 2011. Stimulus Repetition Probability Does Not Affect Repetition Suppression in Macaque Inferior Temporal Cortex. *Cereb Cortex* 21:1547-58
Lichtsteiner P, Posch C, Delbruck T. 2008. A 128x128 120 dB 15 mu s latency asynchronous temporal contrast vision sensor. *Ieee Journal of Solid-State Circuits* 43:566-76
Maguire EA, Frith CD. 2003. Lateral asymmetry in the hippocampal response to the remoteness of autobiographical memories. *J Neurosci* 23:5302-7
Maillard L, Barbeau EJ, Baumann C, Koessler L, Benar C, et al. 2011. From perception to recognition memory: time course and lateralization of neural substrates of word and abstract picture processing. *J Cogn Neurosci* 23:782-800
Mandler JM, Ritchey GH. 1977. Long-term memory for pictures. *J Exp Psychol Hum Learn Mem* 3:386-96
Markram H, Gerstner W, Sjostrom PJ. 2011. A history of spike-timing-dependent plasticity. *Front Syn Neurosci* 3:4
Markram H, Lubke J, Frotscher M, Sakmann B. 1997. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* 275:213-5.
Masquelier T, Guyonneau R, Thorpe SJ. 2008. Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains. *PLoS ONE* 3:e1377
Masquelier T, Guyonneau R, Thorpe SJ. 2009. Competitive STDP-Based Spike Pattern Learning. *Neural Comput* 21:1259-76
Masquelier T, Thorpe SJ. 2007. Unsupervised Learning of Visual Features through Spike Timing Dependent Plasticity. *PLoS Comput Biol* 3:e31
Mo J, Schroeder CE, Ding M. 2011. Attentional modulation of alpha oscillations in macaque inferotemporal cortex. *J Neurosci* 31:878-82

Morrison CM, Conway MA. 2010. First words and first memories. *Cognition* 116:23-32
Moscovitch M, Nadel L, Winocur G, Gilboa A, Rosenbaum RS. 2006. The cognitive neuroscience of remote episodic, semantic and spatial memory. *Curr Op Neurobiol* 16:179-90
Mukamel R, Fried I. 2012. Human intracranial recordings and cognitive neuroscience. *Annu Rev Psychol* 63:511-37
Ohki K, Chung S, Ch'ng YH, Kara P, Reid RC. 2005. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature* 433:597-603
Olshausen BA, Field DJ. 2005. What is the other 85% of V1 doing? In *Problems in Systems Neuroscience*, ed. JL van Hemmen, TJ Sejnowski, Oxford University Press
Pedreira C, Mormann F, Kraskov A, Cerf M, Fried I, et al. 2010. Responses of human medial temporal lobe neurons are modulated by stimulus repetition. *J Neurophysiol* 103:97-107
Penfield W. 1958. *The Excitable Cortex in Conscious Man*. Springfield, Illinois: C.C. Thomas
Penfield W, Perot P. 1963. The brain's record of auditory and visual experience. *Brain* 86:595-696
Perez-Carrasco JA, Zamarreno-Ramos C, Serrano-Gotarredona T, Linares-Barranco B. 2010. On neuromorphic spiking architectures for asynchronous STDP memristive systems. *IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1659-62. Paris
Piolino P, Giffard-Quillon G, Desgranges B, Chetelat G, Baron JC, Eustache F. 2004. Re-experiencing old memories via hippocampus: a PET study of autobiographical memory. *Neuroimage* 22:1371-83
Plaut DC, McClelland JL. 2010. Locating object knowledge in the brain: comment on Bowers's (2009) attempt to revive the grandmother cell hypothesis. *Psychol Rev* 117:284-90
Querlioz D, Bichler O, Gamrat C, Ieee. 2011. *Simulation of a Memristor-Based Spiking Neural Network Immune to Device Variations*. 1775-81 pp.
Quian Quiroga R, Kraskov A, Koch C, Fried I. 2009. Explicit encoding of multimodal percepts by single neurons in the human brain. *Curr Biol* 19:1308-13
Quian Quiroga R, Kreiman G. 2010. Measuring sparseness in the brain: comment on Bowers (2009). *Psychol Rev* 117:291-9
Quian Quiroga R, Kreiman G, Koch C, Fried I. 2008. Sparse but not 'Grandmother-cell' coding in the medial temporal lobe. *Trends Cogn Sci* 12:87-91
Quian Quiroga R, Reddy L, Kreiman G, Koch C, Fried I. 2005. Invariant visual representation by single neurons in the human brain. *Nature* 435:1102-7
Rachmuth G, Shouval HZ, Bear MF, Poon CS. 2011. A biophysically-based neuromorphic model of spike rate- and timing-dependent plasticity. *Proc Natl Acad Sci U S A* 108:E1266-E74
Reddy L, Quiroga RQ, Wilken P, Koch C, Fried I. 2006. A Single-Neuron Correlate of Change Detection and Change Blindness in the Human Medial Temporal Lobe. *Curr Biol* 16:2066-72
Robinson DA. 1968. The electrical properties of metal microelectrodes. *Proc IEEE* 56:1065-71
Rubin DC, Wenzel AE. 1996. One hundred years of forgetting: A quantitative description of retention. *Psychol Rev* 103:734-60
Shepard RN. 1967. Recognition Memory for Words Sentences and Pictures. *J Verb Learn Verb Behav* 6:156-63
Shoham S, O'Connor DH, Segev R. 2006. How silent is the brain: is there a "dark matter" problem in neuroscience? *Journal of Comparative Physiology a-Neuroethology Sensory Neural and Behavioral Physiology* 192:777-84
Standing L. 1973. Learning 10,000 pictures. *Q J Exp Psychol* 25:207-22
Strukov DB, Snider GS, Stewart DR, Williams RS. 2008. The missing memristor found. *Nature* 453:80-3
Swadlow HA. 1988. Efferent neurons and suspected interneurons in binocular visual cortex of the awake rabbit: receptive fields and binocular properties. *J Neurophysiol* 59:1162-87
Thorpe S. 1995. Localized versus distributed representations. In *The Handbook of brain theory and neural nettworks*, ed. MA Arbib, pp. 549-52: MIT Press
Thorpe SJ. 2002. Localized Versus Distributed Representations. In *The Handbook of Brain Theory and Neural Networks*, ed. MA Arbib, pp. 643-5. Cambridge, MA: MIT Press
Thorpe SJ. 2011. Grandmother Cells and Distributed Representations. In *Understanding visual population codes - Towards a common multivariate framework for cell recording and functional imaging*, ed. N Kriegeskorte, G Kreiman. Cambridge: MIT Press
Trachtenberg JT, Chen BE, Knott GW, Feng G, Sanes JR, et al. 2002. Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex. *Nature* 420:788-94.
Tulving E. 1985. How many memory-systems are there? *Am Psychol* 40:385-98
Tulving E. 2002. Episodic memory: From mind to brain. *Annu Rev Psychol* 53:1-25
Viard A, Piolino P, Desgranges B, Chetelat G, Lebreton K, et al. 2007. Hippocampal activation for autobiographical memories over the entire lifetime in healthy aged subjects: An fMRI study. *Cerebral Cortex* 17:2453-67
Viskontas IV, Quiroga RQ, Fried I. 2009. Human medial temporal lobe neurons respond preferentially to personally relevant images. *Proc Natl Acad Sci U S A* 106:21329-34
Wang HM, Hu YG, Tsien JZ. 2006. Molecular and systems mechanisms of memory consolidation and storage. *Prog. Neurobiol.* 79:123-35
Waydo S, Kraskov A, Quian Quiroga R, Fried I, Koch C. 2006. Sparse representation in the human medial temporal lobe. *J Neurosci* 26:10232-4
White KG. 2001. Forgetting functions. *Anim Learn Behav* 29:193-207
Zamarreno-Ramos C, Camunas-Mesa LA, Pérez-Carrasco JA, Masquelier T, Serrano-Gotarredona T, Linares-Barranco B. 2011. On spike-timing-dependent-plasticity, memristive devices, and building a self-learning visual cortex. *Front Neurosci* 5:26