

Are all declarative memories false memories?

J. Kuiper*, Prof. Dr. J.M. Koolhaas†, Prof. Dr. E.A. van der Zee‡

June 30, 2011

Abstract

Memory is the primary constituent of our identity. Unfortunately, memories are not always a true representation of reality. So called False Memories are cases in which memories are distorted or otherwise incongruent with reality. Three cases of False Memory will be reviewed namely source-monitoring errors, suggestibility and bias. Explaining False Memory, however, within the classical framework of memory systems is inadequate, so a different perspective is needed. The Complementary Learning System in combination with Spike Dependent Plasticity, Long Term Potentiation and the philosophical background of Dynamic Systems provides a deeper insight in to the principles of memory and the phenomenon of False Memory.

Keywords: Behavior, Memory, False Memory, Dynamic Systems, Computational Neuroscience

Introduction

Memory is at the core of our identity. All our previous experiences, past and present ideas, learned skills and endeavors are familiarly consolidated within our memories. Devastating diseases like Alzheimer's disease, which primarily act on declarative memory systems, have been adequately described as identity destroying. Indeed, memory can be said to be the primary constituent of identity. Despite this fact, the idea that memories once consolidated remain stable even in healthy individuals has been disputed for many years. The familiar notion of forgetting is something everyone can likely relate to. Recent discoveries show that even remembering can destabilize memories, creating the possibility of retrieval-induced forgetting (MacLeod, 2002; Ciranni & Shimamura, 1999).

Different from the various types of forgetting is that memories themselves can be faulty, e.g. they may represent an event that never took place or details that are attributed to wrong sources. Bartlett (1932) already noted that gaps in perception are filled by previous experiences thus generating a more detailed picture than actually perceived. What we report to have perceived seems often more in line with our expectations than with reality.

Experimentally this was for example shown in a study of Brewer and Treyens (1981). Participants were asked to describe the examiners waiting room after they had participated in some fake experiment, what they recalled is stereotypically associated with an academic office: stacks of books, where no books were actually present.

More controversially under influence of suggestibility false memories seemingly can be created. One detailed account of this phenomenon is the Bijlmer accident, which took place in 1992. An El Al cargo plane crashed into an

eleven story apartment building, killing 43 people. People who were asked 10 months later if they had seen video footage of the airplane crashing into the building reported that they did, although no such footage existed (Crombag et al., 1996).

Under the umbrella term of False Memory, this paper will review the cases in which memories are distorted or otherwise incongruent with reality. Recent understandings of memory systems will be the basis of explaining these cases and memory in general. The classical Atkinson and Shiffrin (1968) model will be the framework in which the various anomalies are initially posed, but it will become apparent that this model has limitations.

The aim of this paper is to propose a computational perspective of memory as an alternative to deal with these limitations and to provide a more general explanation of memories. Drawing from dynamical systems theory, neuroanatomy and molecular neuroscience these underlying thoughts will be elaborated further within this paper and provide an explanation for the phenomenon of false memories. It will be argued that false memories are part of the general mechanisms of memory, and that they occur more often than it seems.

Memory

Types and definition

To properly frame the problem of false memories a definition of memory needs to be given. Recalling a vivid memory like a first kiss or high school graduation seems to be the epitome of our memory system; lively images, sounds and smells can be relived in great detail. Unfortunately, taking those experiences as the definition of memory severely underestimates the amount of recalling that occurs unconsciously (implicitly) or can never reach our conscious experience.

The taxonomy of memory (see figure 1) indeed shows that implicit memory systems make up for large amounts of learning. But the existence of false implicit memories, e.g. false motor skills, has yet to be indicated and their non-verbal nature makes their existence hard to study. This article will therefore focus on the declarative (explicit) memories that are defined to be conscious, or at least capable of entering conscious thought.

Within the declarative memories a sub-division between episodic and semantic memory has often been made. Episodic memories are those that have context, like a 16th birthday, whereas semantic memories are independent of context, knowledge of the capital of Germany, for example.

The division ultimately becomes vague: is a favorite childhood toy an episodic or semantic memory? Moreover,

*University of Groningen, Dept. of Artificial Intelligence

†University of Groningen, Dept. of Behavioral Physiology

‡University of Groningen, Dept. of Molecular Neurobiology

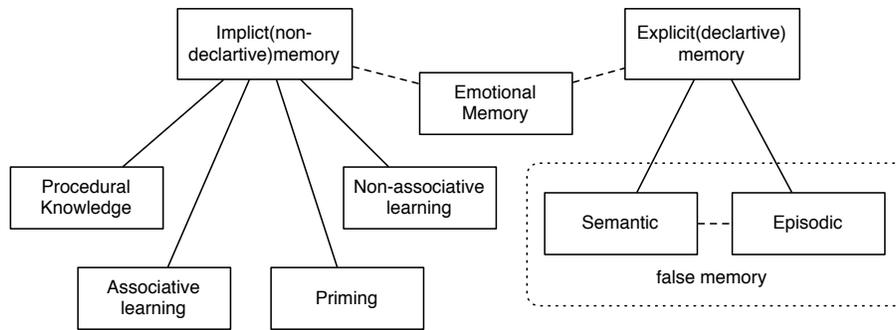


Figure 1. Taxonomy of memory

when a chunk of semantic information is repeatedly presented within a context, like when a friend tells a trivia, over time this chunk loses its source, the friend, and manifests itself as a semantic memory (Tulving, 1985). The idea that information in episodic memories gradually changes into semantic information is even anecdotally present when we try to tell a story to somebody, but that person is ironically the stories' source. So too can it be reasoned that we can change a semantic memory into an episodic one by placing the object or idea within a made-up context, as the case with some mnemonic techniques.

Because large amounts of interactions occur between semantic and episodic memories some neuroscientists have posed that the division is ultimately a trivial one. The idea is that episodic memory might merely be a different level of conscious experience within semantic ones. In view of this discussion the division will be ignored altogether within this article.

Still, the question remains: what exactly is a memory? What differentiates a true memory from, say, a lively fantasy, dream, delusion, hallucination or indeed a false memory. The exact definition of memory is an illusive one but some criteria must be kept in mind.

Firstly memories are acquired (i.e. learned), they come forth from being situated in a world, and are not innate to the brain. Memory is a constructive process. Secondly, the acquired memories are somehow placed within the neural networks of our brain, this neural correlate of memory is commonly referred to as a *trace* or *engram*. Thirdly, memory traces can be reactivated given a certain context, whether it is an external stimulus or an internal narrative, the ability to recall is crucial. It does not mean, however, that declarative memories are recalled only within a certain context (as is the case with associative learning like operant conditioning), but it points to the concept of being "readily available". And finally memory is organized. Not only in temporal and spatial terms but also semantically memories seem to be clustered into meaningful groups (Collins & Loftus, 1975). The meaningful clustering also happens on the more basal level of perceptual binding, input from multiple sensory modalities are stored as a single event.

Defining false memory

There are various types of false memories, not all of them in violation of the criteria described above. A recalled fantasy might be well within the definition of a true memory, as long as it is remembered that it is a fantasy. Therefore the criteria as given by Schacter (1999) in *The Seven Sins of Memory* will be used: misattribution, suggestibility and bias. Underlying all of these phenomena is, however, that the person believes them to be true while they are, in fact, distortions.

Misattributions: Misattributions are instances when memory is attributed to a wrong source. They are therefore also called source-monitoring errors. Three variations of source-monitoring can occur: external, internal and reality (for a review see: Johnson et al., 1993). In case of external misattribution the source of one real event is attributed to another real event. It has been proposed as an explanation for many false allegations due to incorrect eyewitness testimony (Lindsay & Johnson, 1989). When an eyewitness is exposed to various mug shots of people prior to a lineup the witness is likely to pick someone recognized from the mug shots, instead of the actual perpetrator. In other words, the memory of the person in the mug shots is incorrectly attributed to the memory of the crime.

Internal source monitoring is the ability of discriminating between what one thought and between what one said. It is quite possible that one thinks he or she said something but never actually spoke.

Reality monitoring is a more complex phenomenon requiring both errors in internal and external source monitoring. Confabulations are a primary example of reality monitoring errors. People suffering from chronic alcohol-abuse induced Korsakoff or cerebral contusions often confabulate. They create memories of episodes that never happened from past experience and internal narrative, often with a distorted temporal and spatial context. Anticholinergic agents like scopolamine and atropine are also known to cause confabulations. Even healthy individuals might confabulate however, being convinced that some event actually took place, but never did.

Experimentally source-monitoring errors can be seen in

the Deese-Roediger-McDermott (DRM)-paradigm (Deese, 1959; Roediger & McDermott, 1995). Herein participants are shown a list of semantically related words (e.g., NIGHT, BED, DROWSY and DREAM). When, after reading the list, people are asked to name the words in a free-recall task they often report words that are related, but not present in the list (in this case SLEEP). This is a case of internal source monitoring. The DRM-paradigm has therefore often been cited to show that false memories can be created under controllable circumstances.

Suggestibility: different from source monitoring is suggestibility. Exemplary cases of suggestibility are those of leading (or loaded) questions. “Was the man wearing a hat?” immediately suggests that there was a man, even though the person might have been female.

In one experiment showing this effect participants are given a series of pictures of a car waiting for a stop sign. The participants were then divided into two groups and were asked several questions about the pictures. In one of the groups the questions contained misleading information, like questions about a yield sign instead of the stop sign. The control group was given no misleading information about the sign. When the participants from both groups are forced to choose which picture they saw in the initial phase, those who were given misleading information were more likely to pick the incorrect yield sign than the controls (Loftus, 1979).

Bias: the other case of false memory, bias, is where no misleading information is needed. Participants by virtue of their prior beliefs and experiences pick a wrong item in a selection task. Especially emotionally loaded subjects like marriage are susceptible to bias (Karney & Coombs, 2000). People tend to evaluate memories by their present emotional evaluation, not by their past ones leading to incorrect judgment of emotional memories. Emotional bias might for example direct attention away from the correct memory and instead retrieve irrelevant ones.

Bias and suggestibility share with source-monitoring that preexisting knowledge is somehow incorrectly linked to incoming stimuli. This incorrect linkage should be explained within our current understandings of memory systems. False memories are part of our normal memory formation, and should therefore be explained in terms of that process.

A model of false memory

The Atkinson and Shiffrin (1968) model has often been used to describe interactions between various discrete memory components. The model will be briefly discussed and various points where false memories could occur will be elaborated upon.

Figure 2 shows the basics of the model. Although many alterations have been proposed, like the phonological loop in short term memory rehearsal (Baddeley, 1992) and various intermediates between short term memory (STM; also known as working memory) and long-term memory

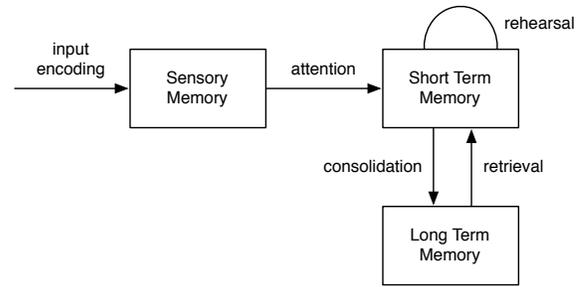


Figure 2. Simplistic model of memory

(LTM), the underlying model has in principle remained unchanged.

The model is based on the idea that the path from sensory input to our long-term memory system is a multi-staged one. Before a memory can be properly consolidated the sensory input must first be encoded into neuronal firing patterns. Passing through our sensory memory it will then, under influence of attention, enter STM. Within the short term memory various rehearsal and selection processes take place. Finally a very limited amount of the initial input will enter our LTM. The step from STM to LTM is a crucial one since memories in LTM are less likely to destabilize. The classical view is that once an item is consolidated into LTM the memory is stable, it has undergone the necessary maturation so it will not be susceptible to interference, such as the presentation of distractors. Recent research, however, disputes this principle (see: Dudai & Eisenberg, 2004). It seems that after a memory is retrieved it has to undergo a process of reconsolidated, making it again susceptible to interference.

Considering the various stages of the model there are several places where false memories might emerge. Firstly, at the level of input encoding. This is known to be subjected to the phenomenon of perceptual closure (or gap completion). A well known example is that of broken drawings (Gollin, 1960). In this experiment participants are shown incomplete drawings of familiar objects. These pictures become progressively more complete, and the participants have to report when they identify the depicted object. With practice participants learn to identify the broken drawings more quickly.

Interestingly, even patients with anterograde amnesia show improvements in their ability to identify the pictures, although they don't recall having seen the images (Milner, 1970). Another example of implicit effects in input encoding is the psychological effect of perceptual priming. Priming (classically called preparing) is the effect of faster response times to perceptually or semantically similar objects (e.g., Pecher et al., 1998) a phenomenon that can also be seen in patients with anterograde amnesia.

Priming and perceptual closure of visual stimuli can be localized in the model (figure 2) at the level of input encoding and sensory memory. However, it is an unlikely candi-

date for explaining the recall of non-existing items in the DRM-paradigm due to its unconscious nature.

Further within the model the recall of non-existing items can be explained in several ways. One possible explanation is that when the stimulus (i.e., the list in the DRM-paradigm) enters STM, similar items are retrieved from the LTM and compared to the stimulus, resulting into incorrect consolidation of the extra retrieved objects. Another possibility is that only the gist of the list is consolidated into LTM, and that when it is retrieved semantically similar words tag along.

Whether False Memories occur during retrieval or consolidation has been a matter of debate. Proponents of the retrieval explanation have pointed out that participants might simply not be bothered with pointing out the mistakes in a forced recognition task. It might also be the case that the effects of suggestibility are temporary, the participants might later realize that their given account was incorrect, and critical thinking might simply be a slower process. Those in favor of the consolidation theory present that false memories can be remarkably stable and that memory is susceptible to amnesia and so too false memory should be susceptible to the same process. This has been backed up by neuroimaging studies showing activity in memory related areas during tasks that included misinformation (Okado & Stark, 2005; but see for a review of neuroimaging studies Schacter & Slotnick, 2004).

In summary, it is hard to pinpoint which part of this model is responsible for false memory: retrieval, consolidation or both. Furthermore where exactly in the intermediate steps from STM to LTM we are susceptible to false memory remains unclear. The discovery of the process of reconsolidation (Riccio, Millin, & Bogart, 2006) opens a window to the identification of new types of false memory. The unifying criterion of all false memories is that preexisting information is somehow incorrectly linked with the new trace or retrieval. Preexisting knowledge seem to be incorporated by the activation of semantically related concepts. The psychological concept of spreading of activation provides an theoretical explanation of how old information might be (falsely) incorporated. However, spreading of activation does not provide a neurological background. On the other hand, imaging studies implicate the neocortical and hippocampal areas as a neural substrate for the memory systems, but do not provide the necessary psychological framework for false memory.

To understand false memory, not only the psychology but also the neuroscientific background of memory formation must be understood. Marr (1982) was the first to introduce the three levels of description necessary to form a comprehensive understanding of mind: the implementation, the algorithms at work and their purpose. This principle will be applied to propose explanation to phenomena of false memory and memory in general.

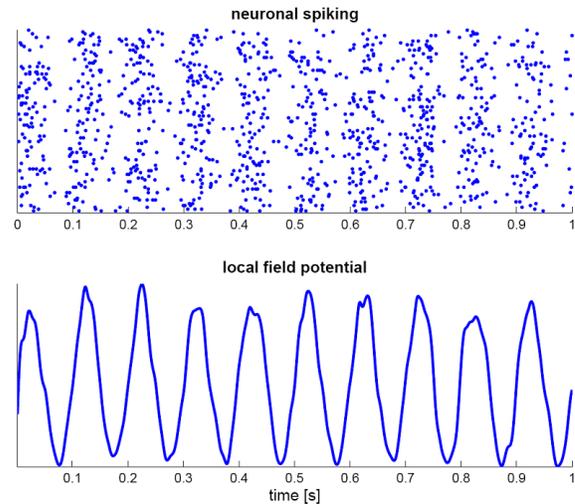


Figure 3. Simulation of neural oscillations at 10 Hz. Upper panel shows spiking of individual neurons (with each dot representing an individual action potential within the population of neurons), and the lower panel the local field potential reflecting their summed activity (Wikipedia, 2011)

Contextualized memory systems

Causality & Synchronicity

The principle of coincidence detection plays an important role in memory formation. When something happens in the real world, input from different sensory modalities (such as visual and auditory) is encoded into neural firing patterns. These neural firing patterns follow a path within the neural network of our brain, and at some points synchronous neural activity from various sensory modalities exceeds the firing thresholds of connected neurons. Those neurons that respond to that precisely timed arrival can be said to detect the coincidence between the two independent firing patterns. When, for example, the activity from the auditory nerve after a loud bang coincides with the optical nerve activity from a flash the activity of the neurons that respond to that coincidence bind together the two firing patterns into one stream. The synchronous activity from the bang and the flash thus encodes for a semantically meaningful event, for example that of fireworks. Because the outside world by virtue of the laws of physics follows a predictable pattern, the synchronous stream caused by the coincidence can be said to encode for real physical dynamics. Synchronous neural activity provides a method for encoding for such coincidences because it seems to be able to bind together multiple sensor modalities.

This synchronous activity can be demonstrated with EEG (see figure 3). Various studies have shown Event Related Potentials (ERPs) in association areas of the cortex, showing high levels of local synchronicity after a perceived input or thought (Singer & Gray, 1995). Gamma band oscillations (30 – 100 Hz) have often been implicated in the binding of sensory modalities as in the example of the fireworks.

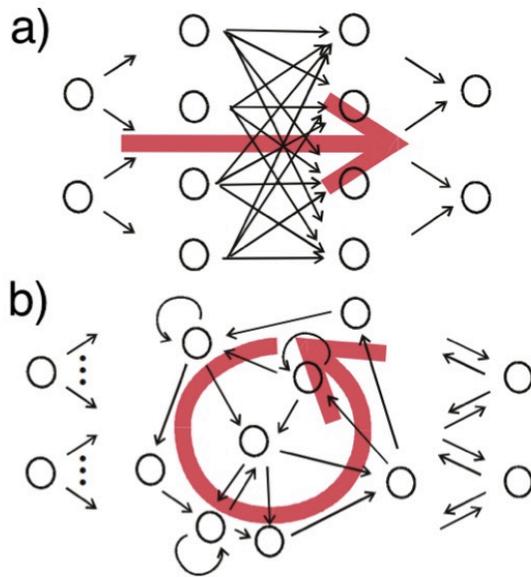


Figure 4. Two common types of neural network graphs. The network a) illustrates a feedforward pattern. Notice the ordering of layers. Network b) illustrates a typical recurrent network, such as a Hopfield (1982) network. (reproduced from Jaeger (2001))

Furthermore, gamma and theta (4 – 7 Hz) band oscillations have been found in the hippocampus and entorhinal cortex during memory formation and retrieval (Klimesch et al., 2010; Axmacher et al., 2006).

However, while binding multiple sensor modalities might encode for objects or singular events in the world, often more complex causal relations have to be inferred. The perception of the world happens within a certain temporal context; things happen in a specific order and being able to infer those dynamics is crucial for contextual adaptation. For example, when a glass falls from a table onto a concrete floor it is empirically acceptable and useful to predict that it will break. To account for these causal temporal dynamics high levels of synchronicity are not sufficient.

If highly synchronous activity happened without any form of temporal organization it's representative power would be severely limited. Thus delays within the connections between the individual neural ensembles are necessary for the precisely timed spike train intervals to encode for real time dynamics.

Neural topology

Cellular imaging techniques have allowed recording of such real time dynamics and mapping the connections between neurons, the resulting relationship has been named the connectome (Bullmore & Sporns, 2009). Two principles seem to be important: feedforward and recurrent (see figure 4) dynamic activity.

Computational modeling of these two principles show that feedforward connections might introduce delays by

virtue of speed of neural transmission, sub-threshold membrane dynamics and various other mechanisms. Recurrent connections, however allow precisely timed spikes to be generated in an oscillating manner.

Mathematical descriptions of recurrent networks that are trained on certain input patterns by modeling Hebbian-learning, such as Hopfield (1982) networks have shown to be capable of pattern completion and noise reduction with multiple stable patterns. More recent computational advances have also shown that sparsely distributed recurrent networks, such as Echo State Neural Networks (Jaeger, 2007) and Liquid State Machines (Maass, Natschl, & Markram, 2003), can be computationally trained to mimic complex nonlinear harmonics. These complex nonlinear harmonics can be interpreted as a mathematical description of real world causality, thus the ability to mimic it closely resembles the ability of memory.

Recurrent networks allow for a fast type of memory by keeping patterns of synchronous activity locked in various timed loops. They can echo causal relations of the outside world by virtue of delays in synaptic transmission as well as their network organization. It should be noted that there is no need for any single neuron in these networks to be associated with only one trace of synchronous activity. Neither should any one trace be encoded in only one network. It seems possible that there are many different traces of synchronous activity within the connectome for a single causal relation, and new experiences (i.e. novel encoded sensor input) do not necessarily interfere with old traces. Indeed, neural networks are a highly distributed form of encoding and remembering sensory input.

Models of recurrent networks can thus encode for the causal temporal dynamics of the world. And precisely timed synchronicity between spike trains (Kumar, Rotter, & Aertsen, 2010) in neural networks seems to be capable of this by binding multiple traces together, such as those from multiple sensor modalities. Reminiscent activity in the recurrent connections seems to allow for memory and a background for pattern completion. In summary, precise timing of the delays in the networks create a representation, an echo, of the dynamics of the outside world: a mental representation of physical coincidence.

Remembering and future prediction

Causal relations have a predictive nature, for example the falling glass predicts that if a glass falls it might break. So, memories of physical causality can be used to predict future events. In this case input from the optical nerve coincides with reminiscent activity in recurring neural networks generating activity that completes the pattern of a breaking glass. By virtue of the dynamics of the outside world recurrent neural networks and precisely timed spike train synchronicity are not only inherently memory systems but also a future prediction systems. Thoughts about the future and remembering might actually be two sides of the proverbial coin. This pattern of this memory and future prediction

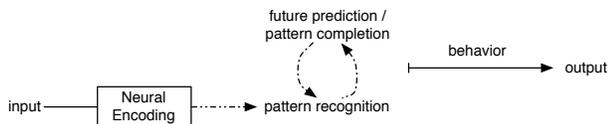


Figure 5. Conceptualization of memory formation. Dotted lines represent spike trains, solid lines conceptual connections.

loop are very schematically presented in figure 5.

False coincidences might trigger unrelated patterns and hence false memories. However, explaining false memory in terms of high synchronicity is limited because this is not sufficient to explain long term stabilization. It is false thought or false working memory at best. In general it can be said that the precise timing of action potentials in the connectome creates an echo of the worlds' dynamics. But (false) memories are more than a complex echo of coincidences, they have a bias for common patterns. This bias for common patterns memories also allows us to extract relevant information from chaotic input and complete more abstract (predictive) patterns.

The Hebbian-paradigm has often been used to explain this bias for common patterns. Because neuronal connections strengthen or weaken on synchronous activity (Spike Time Dependent Plasticity, STDP), long-term memory effects can occur due to the slight adjustments in the precise timing (e.g. Kozloski & Cecchi, 2010). Interference with recent memories is typically greater than that on remotely acquired memories, because connections have been strengthened over time (Ribot's law, also called the Law of regression). This process of creating a bias for certain patterns due to the slower effects of Hebbian-learning has been called consolidation. On a cellular level consolidation can be seen in Long Term Potentiation (LTP) and Long Term Depression (LTD)(see for a review: Malenka & Bear, 2004). False memories might occur due to the improper favoring by STDP of a previous trace to a incoming stimuli. I.e. when incoming stimuli falsely triggers a pattern of activity that pattern might stabilize due to the plasticity, such that when the pattern is recalled the improper pattern is also again falsely activated thus being a false memory.

Long term false memory

When information is improperly linked to incoming stimuli (either external or internal), this does not necessarily constitute a false memory. It is still a matter of debate whether memories ever truly stabilize, however data from amnesiacs reveal that there is an organizational effect in memory. Lesions or epilepsy in the Medial Temporal Lobe (MTL, which includes the hippocampal region and the adjacent perirhinal, entorhinal, and parahippocampal cortices) can cause severe forms of amnesia. The classical cases in which the MTL was bilaterally removed resulted in almost pure

anterograde amnesia. High levels of protein activity linked to LTP and LTD can be found in the hippocampal areas. Both imply a role of the MTL in the creation of false memories.

The view that interference of new input with recently formed memory occurs is also supported by studies on the role of the hippocampus in memory formation. The hippocampal area is a fast system for memory formation. It provides links to older traces and incoming information from the association areas (through the entorhinal cortex) in the neocortex allowing to complete patterns based on existing traces. Gradually these links cause reorganization in the neocortical areas in such a way that the trace becomes independent of the linkage from the hippocampal areas. This neocortical-hippocampal system of memory formation has been called the Complementary Learning System, (for a review see Simons & Spiers, 2003; Ron Sun, 2008)

To minimize interference between non-overlapping episodes the hippocampus has a tendency to assign separated patterns to episodes in the hippocampal CA3-subfield.

This pattern-separation is also greatly facilitated by strong inhibition from the Dentate Gyrus (DG). The DG has almost no overlapping representations and to ensure a consistency, new neurons are constantly added through neurogenesis.

But such pattern-separated representation has little generalization power and is, even with the inhibitory connections, susceptible to catastrophic interference. Catastrophic interference in this context means that incoming information might destabilize most of the existing representations. The neocortex allows for more stable representations by being a slower learning system (McClelland et al., 1995; Squire & Zola-Morgan, 1991; Wiltgen et al., 2004). These complementary systems allow real world dynamics to be echoed (cortical association areas), tagged and pattern completed (CA1 and CA3 subfields) and then generalized into new abstract patterns (neocortex).

What seems to be happening at a more meta-level was postulated by the Parallel Distributed Processing paradigm (Rumelhart & McClelland, 1986a, 1986b). Gradual changes in the parallel distributed processing of the brain allows for pattern completion and future prediction. These gradual changes happen in the direction of the perceived input, that is our inner mental representation of the dynamics of the world becomes more similar to what we perceive. These gradual changes happen in the direction of the perceived input, i.e. our inner mental representation of the dynamics of the world becomes more similar to what we perceive. A formal model of this gradual updating is known as the Temporal Context Model (TCM) (Howard M.W. & Kahana M.J., 2002; Howard et al., 2005). Inputs that are similar to the *inner context* are more likely to activate and gradually update that *inner context* in the direction of that input.

TCM is a form of mental drifting that explains why more

recent things are easier to remember than older things. This effect is known as the primacy and recency effect, and TCM has been successfully applied in explaining these effects in a free recall task. It also hypothesizes that episodes that are very different from previous experience are harder to learn and cause more significant updating (reorganization).

Are all declarative memories false?

In summary, the world behaves according to specific causal dynamics. These causal dynamics, coincidences, are encoded into specific traces of synchronous activity which due to recurrent connections form reminiscent background activity, like an echo. When new sensory input arrives this is overlaid on top of the background activity, filtering input that fails to exceed thresholds but also completes patterns. This pattern completion not only happens on a semantic level (e.g. completing broken pictures) but also on a temporal (causal) level, allowing inferences about the future to be made.

Due to the precise timing of spike trains in these traces some connections are strengthened or weakened due to the effects of Spike Time Dependent Plasticity. This strengthening and weakening happens in a gradual manner, slowly updating the background activity to represent that of the outside world. This process inherently favors coincidences that occur often, not only in the real world but also those that coincide with the background activity. It allows for a more robust way of completing patterns (spreading of activation) and predicting future.

This theme means parting from the idea that memories are somehow stabilized and stored away for later use. Instead, they are an integrate part of our (conscious) thought and far more dynamic and versatile. The notion of consolidation might therefore be a misnomer, memories might actually never truly stabilize but are constantly adjusted. Indeed the neural activity in our brain is both memory (consolidation) and thought (retrieval) at the same time, mimicking input from the real world. It allows for vast pattern completion and both spatial and temporal reasoning, for example inferring that cups that fall might break.

As a consequence some patterns might be incorrectly completed with previous experience. This false synchronous activity might go unnoticed, and if the coincidence is strong enough to stay remnant in the background activity it can then over time become stable due to the strengthening of the connections through spike time dependent plasticity. The stabilization of such false synchronous activity might actually happen quite often, but go unnoticed. Only when a behavioral response such as a verbal report is out of sync with what is *known* about the world these false memories might get noticed. Neural activity only *represents* the dynamics of the world in its precisely timed spike trains, so all memories could in principle be labeled as false memories. And, discrepancies with the

real world might indeed go forever unnoticed.

It can be reasoned that the existence of non-declarative false memories is therefore unlikely since a discrepancy between the remembered dynamics and actual dynamics of the physical world can cause great harm. Subjects suffering from cerebellar damage indeed can show deficient fine motor skills. It seems there is more feedback from the surroundings in a faster pace than with declarative memories, and thus a reduced likelihood in persistent false memories.

However the distributed form of memory consolidation allows for very dynamic updating which is of importance in an ever changing world. So too the ability to readily make associations and infer causal relations is crucial. It can be speculated that storing memories in terms of adjusting global activation to mimic outside activity greatly reduces the adjustments (and thus energy) needed to remember new things. It is a very efficient way of storing new data. Indeed, the completion of patterns and predicting the future is of such importance for contextual adaptation, and thus survival, that false memories are a small price to pay.

Acknowledgements

van der Zwan, J. E. for her initial work on the topic within her bachelor thesis.

References

- Atkinson, R., & Shiffrin, R. (1968). *The psychology of learning and motivation (Volume 2)*. New York: Academic Press.
- Axmacher, N., Mormann, F., Fernández, G., Elger, C. E., & Fell, J. (2006, August). Memory formation by neuronal synchronization. *Brain research reviews*, 52(1), 170–82.
- Baddeley, A. (1992, January). Working memory. *Science*, 255(5044), 556–559.
- Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press.
- Brewer, W. E., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, 13(2), 207–230.
- Bullmore, E., & Sporns, O. (2009, March). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews. Neuroscience*, 10(3), 186–98.
- Ciranni, M. A., & Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of experimental psychology Learning memory and cognition*, 25(6), 1403–1414.
- Collins, A., & Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychological review*, 82(6).
- Crombag, H., Wagenaar, W., & Van Koppen, P. (1996). Crashing memories and the problem of 'source monitoring'. *Applied Cognitive Psychology*, 8, 95–105.
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, 58(1), 17–22.
- Dudai, Y., & Eisenberg, M. (2004, September). Rites of

- passage of the engram: reconsolidation and the lingering consolidation hypothesis. *Neuron*, 44(1), 93–100.
- Gollin, E. (1960). Developmental studies of visual recognition of incomplete objects. *Perceptual and Motor Skills*.
- Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79(8), 2554.
- Howard, M. W., Fotedar, M. S., Datey, A. V., & Hasselmo, M. E. (2005, January). The temporal context model in spatial navigation and relational learning: toward a common explanation of medial temporal lobe function across domains. *Psychological review*, 112(1), 75–116.
- Howard M.W., & Kahana M.J. (2002). A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology*, 46(3), 31.
- Jaeger, H. (2001). *The "echo state" approach to analysing and training recurrent neural networks* (No. GMD Report 148). GMD Forschungszentrum Informationstechnik, Sankt Augustin.
- Jaeger, H. (2007). *Echo state network* (Vol. 2) (No. 9).
- Johnson, M., Hashtroudi, S., & Lindsay, D. (1993). Source Monitoring. *Psychological Bulletin*, 114, 3–3.
- Karney, B. R., & Coombs, R. H. (2000, October). Memory Bias in Long-Term Close Relationships: Consistency or Improvement? *Personality and Social Psychology Bulletin*, 26(8), 959–970.
- Klimesch, W., Freunberger, R., & Sauseng, P. (2010, June). Oscillatory mechanisms of process binding in memory. *Neuroscience and biobehavioral reviews*, 34(7), 1002–14.
- Kozloski, J., & Cecchi, G. a. (2010, January). A theory of loop formation and elimination by spike timing-dependent plasticity. *Frontiers in neural circuits*, 4(March), 7.
- Kumar, A., Rotter, S., & Aertsen, A. (2010, September). Spiking activity propagation in neuronal networks: reconciling different perspectives on neural coding. *Nature Reviews Neuroscience*, 11(9), 615–627.
- Lindsay, D., & Johnson, M. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, 17(3), 349–358.
- Loftus, E. (1979). *Eyewitness Testimony*. Harvard University Press.
- Maass, W., Natschl, T., & Markram, H. (2003). Computational Models for Generic Cortical Microcircuits. *Computational neuroscience*, 1–26.
- MacLeod, M. (2002). Retrieval-induced forgetting in eyewitness memory: forgetting as a consequence of remembering. *Applied Cognitive Psychology*, 16(2), 135–149.
- Malenka, R. C., & Bear, M. F. (2004, September). LTP and LTD: an embarrassment of riches. *Neuron*, 44(1), 5–21.
- Marr, D. (1982, June). Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995, July). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, 102(3), 419–57.
- Milner, B. (1970). Memory and the medial temporal regions of the brain. *Biology of memory*.
- Nalbantian, S., Matthews, P. M., & McClelland, J. L. (2010). *The Memory Process: Neuroscientific and Humanistic Perspectives*. The MIT Press.
- Okado, Y., & Stark, C. E. L. (2005). Neural activity during encoding predicts false memories created by misinformation. *Learning Memory*, 12(1), 3–11.
- Pecher, D., Zeelenberg, R., & Raaijmakers, J. G. W. (1998). Does Pizza Prime Coin? Perceptual Priming in Lexical Decision and Pronunciation. *Journal of Memory and Language*, 38(4), 401–418.
- Riccio, D. C., Millin, P. M., & Bogart, A. R. (2006). Reconsolidation: a brief history, a retrieval view, and some recent issues. *Learning & memory (Cold Spring Harbor, N.Y.)*, 13(5), 536–44.
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal Of Experimental Psychology. Learning Memory And Cognition*, 21(4), 803–814.
- Ron Sun. (2008). *The Cambridge Handbook of Computational Psychology (Cambridge Handbooks in Psychology)*. Cambridge University Press.
- Rumelhart, D. E., & McClelland, J. L. (1986a). *Parallel Distributed Processing* (Vol. 1; D. E. Rumelhart & J. L. McClelland, Eds.). MIT Press.
- Rumelhart, D. E., & McClelland, J. L. (1986b). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 1; D. E. Rumelhart & J. L. McClelland, Eds.) (No. 2). MIT Press.
- Schacter, D. L. (1999). The Seven Sins of Memory. *American Psychologist*, 54(3), 182–203.
- Schacter, D. L., & Slotnick, S. D. (2004, September). The cognitive neuroscience of memory distortion. *Neuron*, 44(1), 149–60.
- Simons, J. S., & Spiers, H. J. (2003, August). Prefrontal and medial temporal lobe interactions in long-term memory. *Nature reviews. Neuroscience*, 4(8), 637–48.
- Singer, W., & Gray, C. M. (1995, January). Visual feature integration and the temporal correlation hypothesis. *Annual review of neuroscience*, 18, 555–86.
- Squire, L. R., & Zola-Morgan, S. (1991, September). The medial temporal lobe memory system. *Science (New York, N.Y.)*, 253(5026), 1380–6.
- Tulving, E. (1985). How many memory systems are there? *American Psychologist*, 40(4), 385–398.
- Wikipedia. (2011). *Neural oscillation* — *Wikipedia, the free encyclopedia*. Available from http://en.wikipedia.org/wiki/Neural_oscillations ([Online; accessed 22-Januari-2011])
- Wiltgen, B. J., Brown, R. a. M., Talton, L. E., & Silva, A. J.

(2004, September). New circuits for old memories: the role of the neocortex in consolidation. *Neuron*, 44(1), 101–8.