Long-term memory disorders: measurement and modeling
Meeter, M.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
A connectionist model is presented, the TraceLink model, that implements an autonomous 'off-line' consolidation process. The model consists of three subsystems: (1) a trace system (neocortex), (2) a link system (hippocampus and other areas), and (3) a modulatory system (basal forebrain and other areas). The model is able to account for many of the characteristics of anterograde and retrograde amnesia, including Ribot gradients, transient global amnesia, patterns of shrinkage of retrograde amnesia, correlations between anterograde and retrograde amnesia or the absence thereof (e.g., in isolated retrograde amnesia). In addition, it produces normal forgetting curves and can exhibit permastore. It also offers an explanation for the advantages of learning under high arousal for long-term retention.

5.1 THE TRACELINK MODEL

In chapter 2 of this thesis, we reviewed the principal findings from amnesia. These can, was promised in the introductory first chapter, be explained with help of the TraceLink model of memory. In this chapter, we will attempt to do so. In the subsequent chapter, TraceLink will be applied to semantic dementia. Thereafter, work will be presented on how consolidation is implemented in TraceLink. Part II will end with a discussion of how consolidation could be implemented in the brain, and how consolidation as a theory of remote memory and amnesia compares to alternative hypotheses.

Structure of the model

The connectionist model we propose here incorporates the basic framework presented in the introduction as a possible, neuroanatomically informed solution to the three dilemmas discussed there, the forgetting paradox, the connectivity problem and the stability-plasticity dilemma. A schematic drawing of the model is shown in Figure 6. Its three main components are (1) a trace system, (2) a link system, and (3) a modulatory system.

(1) Normally, the greater part of a memory trace will be stored in the connections of the trace system. The trace system represents roughly the neocortical basis of memories. The input to the trace system originates in sensory areas. We assume considerable preprocessing in these areas, which themselves do not form part of the model. Similarly, output or motor areas are not included, but are rather assumed. We identify the trace system with association areas in the neocortex that have been associated with memory,
such as the temporal lobe neocortex (Miyashita, 1993) and posterior parietal cortex (Izquierdo et al., 1997).

(2) It is the link-system's function to connect remote trace elements (i.e., those without direct cortico-cortical connections). The link system has a much smaller number of elements than the trace system. Among others, this implies that link elements are more likely to be reassigned from an old representation to a new, thus causing interference by new learning. Link-elements are connected to each other and to a random subset of the trace-elements. Connections involving link elements are much more plastic than trace-to-trace connections. This is illustrated in Figure 2 through the close attachment of the link system to the modulatory system. The link system can be identified with the hippocampus and other medial temporal lobe structures, such as the entorhinal and perirhinal cortex.

(3) Activation of the modulatory system causes increased plasticity of the link system. The modulatory system may be activated through central states such as arousal and attention. These are not implemented in this version of the model, but can be mimicked by simply designating certain stimuli as 'interesting' and manually increasing the activation of the modulatory system whenever they are present. As has been shown elsewhere, there is at least one plasticity-affecting state that can be derived directly from the link system: novelty-related arousal (Grossberg, 1976; Murre, 1992). The modulatory system can be identified with the basal forebrain and other subcortical centers that control plasticity in the brain. A structure that certainly also plays a role in the modulatory system is the hippocampus. Many studies have implicated it in novelty processing (e.g., Johnson & Moberg, 1980; Knight & Nakada, 1998; Montag-Sallaz, Welzl, Kuhl, Montag, & Schachner, 1999; Mumby, Gaskin, Glenn, Schramek, & Lehmann, 2002; Zhu, McCabe, Aggleton, & Brown, 1997), and in several models it plays a role in modulating its own plasticity (see chapter 10). The hippocampus thus plays a double role in the TraceLink model: it is part of the link system, but it is also involved in regulating its own plasticity.

Figure 6: Overview of the TraceLink model, showing the link system, the trace system, and the modulatory system (indicated by a ΔW sign, symbolizing control of learning rate on the connection weights in the system). Only a few nodes and a few connections have been drawn in order to prevent clutter.
In the following section we shall illustrate with connectionist simulations how the TraceLink model may account for normal episodic learning and recall, retrograde amnesia and gradients in retrograde amnesia, shrinkage or recovery of retrograde amnesia, transient global amnesia, anterograde amnesia, and for implicit memory.

Details of the connectionist model

The model used in the simulations consists of two components, the trace system and the link system. Since none of the simulations depends on more than a relatively coarse notion of modulation, we made no attempt to model the third component of the TraceLink model, the modulatory system. Instead, its functions are here assumed (but see chapter 10). The trace system is modeled as a layer of 200 nodes, the link system as a layer of 42 nodes. Both layers have internal connections, and are connected with one-another. Every two nodes can in principle be connected. As is the case with a majority of neurons in the cortex, nodes only have excitatory synapses. The nodes model groups of neurons that are at some distance from each other.

Both layers have binary stochastic nodes, i.e. they are 'on' or 'off' with a certain likelihood. The likelihood of firing depends on the balance between the excitatory input from other nodes and inhibition. The excitatory input to a node is the weighted sum of the activation of all nodes connected to it. From this excitatory input, inhibition is then subtracted. Inhibition is constantly fine-tuned so as to keep the average number of active cells in a layer as close to a preset number \( k \) as possible. The inhibition mechanism models the working of inhibitory neurons, which may have as one function to keep activity in their region within certain bounds (Braitenberg & Schüz, 1991). In the model, inhibition is the sum of two variables, fast inhibition \( \tau \) and slow inhibition \( \tau \). Fast inhibition is multiplied by the activation level in a layer, and reflects the activation of inhibitory neurons by the excitatory neurons. Slow inhibition reflects the autonomous activity of inhibitory cells. Both slow inhibition and fast inhibition are increased when too many nodes are active (i.e., more than \( k \)), and lowered when there are too few active nodes (i.e., less than \( k \)). Fast inhibition reacts rapidly to a departure from equilibrium, and slow inhibition moves gradually to establish the equilibrium anew. The number \( k \) is set separately for every layer, and consequently inhibition is regulated separately in every layer.

The weights of excitatory connections are modifiable. They are the locus of learning. The rule used for learning is a variant of Hebb's rule (Hebb, 1949; Singer, 1990) that allows for learning as well as unlearning as a function of contingent and non-contingent activity of the nodes. This rule is generally considered biologically more realistic than others (e.g. the backpropagation rule), since long-term potentiation (Bliss & Lomo, 1973) instantiates a Hebb rule. Weights can vary between 0 and 1, and are changed during learning in a simple, linear fashion. The activation, inhibition, and learning rules are explained in detail in the appendix to this chapter.

Learning is not equally fast for all connections. The learning rate, the rate at which changes are made to the weights during learning, is much lower for the within-trace connections than for the connections within the link layer, or between the link layer and the trace layer; connections between the trace and link layers and within the link layer have equal learning rates. Since nodes model groups of neurons, the connection between two nodes is a function of both the number of synapses between the groups of neurons, and their average strength. Though the hippocampus is known to be unusually plastic (Lopes da Silva, Witter, Boeijinga, & Lohman, 1990), the higher learning rate in the
connections involving the link layer is primarily intended to model the higher connectivity in the regions of the link system (Treves & Rolls, 1994). A higher number of synapses that can be recruited for learning a pattern means that this pattern can be learned faster. This is implemented in the model as the higher learning rate of the connections involving the link layer.

A pattern in the learning set consists of a group of trace nodes and a group of link nodes, which are activated when a pattern is presented. Patterns can overlap in both the trace and the link layer. The number of nodes in the trace layer and the link layer that belong to a pattern is equal to the number of nodes active in the layer in equilibrium ($k$). Because patterns are chosen independently of each other, every two patterns share on average a $k/m$ proportion of their nodes in a layer, where $m$ is the number of nodes in the layer. In both layers $k$ is relatively low compared to $m$, which has two distinct consequences: firstly, that at any time in the simulations only a few nodes of both layers are active, and secondly, that the patterns are sparsely coded with little overlap. Because the hippocampal system is much smaller than the neocortex, the overlap of patterns is greater in the link system than in the trace system.\(^1\)

The values of what one could call ‘equilibrium parameters’ are given in the appendix to this chapter. The function of these parameters is to keep the number of active nodes as close as possible to the equilibrium value $k$, and to prevent wild swings in the number of active nodes from one iteration to the next. These parameters only influence the results in the sense that they make meaningful results possible, they do not have an effect on the pattern of findings. It is often difficult to decide what counts as a parameter in connectionist models, and even more difficult to ascertain how many parameters should count as free in the traditional sense. Excluding the ‘equilibrium parameters’, we are left with a few parameters that can perhaps count as free: the learning rates, and the proportion $k$ of active nodes in equilibrium in both layers relative to the size of the layers (see Table 5).

The main findings in the simulations were quite robust, with the form of most functions staying the same with many different sets of parameters. One set of parameter values, those given in Table 5 was used in all simulations.

5.2 NORMAL LEARNING AND RECALL

Figure 7 illustrates at a conceptual level the formation of an episodic memory trace under normal circumstances. For the sake of exposition, four stages can be distinguished (these do not have theoretical value):

Stage 1. Through the sensory and motor channels a set of trace nodes is activated (filled circles). This represents an episode to be remembered. The activated trace nodes have no direct trace-to-trace connections to each other but they are connected to a number of link

---

\(^1\) The link system has denser representations in our simulations than the trace system. Some fields of the hippocampus use sparse representations, which would seem to contradict our model. TraceLink is not detailed enough, however, to compare the model with details of hippocampal anatomy. The denser representations in the link system expresses the widespread idea that the hippocampus has a smaller capacity than the neocortex (e.g., McClelland, McNaughton, & O'Reilly, 1995).
nodes (only two are drawn). We assume that the mechanisms for local inhibition, outlined previously, keep the number of active nodes small.

Stage 2. The trace nodes activate a set of link nodes. We assume that, also here, inhibitory processes suppress the weakest activated link nodes. The modulatory system becomes activated (darkness of shading indicates activation level) and the learning rate increases. As a result of the increased plasticity, connections between link and trace nodes are strengthened (shown by a thickening of the connections). This can take place in minutes or seconds.

Stage 3. Prolonged or repeated activation of the memory trace through the link system will lead to the gradual formation of trace-to-trace connections. With subsequent reactivations, the modulatory system responds with less and less activation – it gradually habituates to the pattern.

Stage 4. Direct trace-to-trace connections have become very strong. Link-trace connections have either decayed or have been reassigned to other memory traces. The modulatory system shows little reaction to the pattern. In short, the memory trace has become independent of the link system.

The formation of long-range trace connections is not by direct synaptic contact. The possibility of long-range axonal sprouting should not be excluded, but we think it more likely that these connections are established via chains of neurons, as is outlined, for example, by Abeles (1991). Because single synapses are rarely strong enough to drive the post-synaptic neuron (Abeles, 1991), there will be many neurons involved in establishing such long-range connections. Establishing a reliable connection will take repeated exposures to the desired connection pattern (i.e., activation of neurons or neural groups A and B, to be connected over a long-range).

Table 5: Parameters in the presented simulations with the TraceLink model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Trace</th>
<th>Link and Layers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Learning rate during acquisition</strong></td>
<td>0.06</td>
<td>0.4</td>
</tr>
<tr>
<td>* within trace</td>
<td></td>
<td></td>
</tr>
<tr>
<td>* within link and between layers</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Learning rate during consolidation</strong></td>
<td>0.0025</td>
<td>0</td>
</tr>
<tr>
<td>* within trace</td>
<td></td>
<td></td>
</tr>
<tr>
<td>* within link and between layers</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total learning during one acquisition</strong></td>
<td>0.06</td>
<td>0.4</td>
</tr>
<tr>
<td>* within trace</td>
<td>(0.06 * 1 iteration)</td>
<td></td>
</tr>
<tr>
<td>* within link and between layers</td>
<td>(0.4 * 1 iteration)</td>
<td></td>
</tr>
<tr>
<td><strong>Total learning during one consolidation period</strong></td>
<td>0.06</td>
<td>0</td>
</tr>
<tr>
<td>* within trace</td>
<td>(0.0025 * 3 trials of 8 iterations)</td>
<td></td>
</tr>
<tr>
<td>* within link and between layers</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Unlearning rate</strong></td>
<td>75% of the learning rate</td>
<td></td>
</tr>
<tr>
<td>* in all connections</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Number of nodes</strong></td>
<td>200</td>
<td>42</td>
</tr>
<tr>
<td>* in trace</td>
<td></td>
<td></td>
</tr>
<tr>
<td>* in link</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Number of nodes active in equilibrium k</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* in trace</td>
<td>10</td>
<td>(1/20th of the nodes)</td>
</tr>
<tr>
<td>* in link</td>
<td>7</td>
<td>(1/6th of the nodes)</td>
</tr>
</tbody>
</table>
We propose that consolidation of memories is the transformation of stage 2 memories to stage 4 memories through repeated reactivation via the link system (this proposal is similar to the ones of Alvarez & Squire, 1994; McClelland et al., 1995, and others). The speed of this transformation may vary with type of material and with many other factors. For some memories, consolidation may take a very long time (up to several decades). It is at this moment not possible to specify what process underlies consolidation, whether conscious, explicit rehearsal, or also unconscious, more automatic processes are involved. We shall return to this important issue in chapter 8.

Recall is modeled as the retrieval of the whole Trace pattern when part of the Trace pattern is offered as a cue. In the case of a stage 2 memory, the partial cue will typically activate the Link nodes associated with the pattern, which in turn will activate the rest of the Trace pattern. In the case of a stage 4 memory, the cue will be able to activate the rest of the trace pattern directly through strong trace-trace connections. Depending on the amount of involvement of the link system, new learning may occur during recall, but this was not implemented in the present version of the model.
Simulation 1: Normal learning and recall

Method
We first simulated the normal workings of the model: normal learning, normal consolidation, and normal recall. This simulation also served as a control for our simulations of different amnesic states.

In the simulations, the model went through two distinct phases, a learning phase and a test phase (see Figure 8). The learning phase consisted of two alternating sub-phases, acquisition of a pattern and consolidation. During acquisition, the model learned one new pattern. This was followed by a period of consolidation, after which another pattern was acquired. In a simulation, the model learned a list of usually 15 patterns with interspersed consolidation periods, after which all the patterns were tested. We replicated each simulation 200 times.

Patterns consisted of 10 activated trace nodes and 7 activated link nodes. This means that a pattern involved $1/20$ of all trace nodes, and $1/6$ of all link nodes (both proportions are determined by the parameter $k$ of the respective layers). In both layers, the patterns were random sets of nodes, and therefore every two patterns shared on average $1/6$ of their link nodes and $1/20$ of their trace nodes. Each pattern was learned during one iteration with the learning parameter listed in Table 5 (this is computationally equivalent to learning the pattern with a lower learning parameter over more iterations).

After the acquisition of a pattern, the model entered a period of consolidation. In a consolidation period, three consolidation trials occurred (see Figure 8). A single consolidation trial proceeded as follows: The model was allowed to cycle freely for a fixed number of iterations (i.e., 150), and whichever pattern was active at the last iteration was consolidated. The dynamics of the model thus selected a pattern to consolidate, with attractors surfacing that were strong in the combined link and trace system. Consolidation was done for 8 iterations at a low learning rate (see Table 5). The model was allowed to cycle during the consolidation, and, as the activation would wander a little, learning for one iteration with a high learning rate was not equivalent to learning for a few iterations with a low learning rate.

| Acquisition of 1 pattern | Cons. trial | Cons. trial | Cons. trial | Acquisition of 1 pattern | Cons. trial | TEST |
|-------------------------|-------------|-------------|-------------|-------------------------|-------------|
| Acquisition period      | Consolidation period | Acquisition period |                      |                         |              |

**Figure 8:** Diagram showing the order of events in most simulations. A simulation was divided into a learning phase and a test phase. The learning phase was subdivided into alternating acquisition periods, in which one pattern was acquired, and consolidation periods. Consolidation periods consisted of three consolidation trials.
The model usually did not settle on the same pattern for all three consolidation trials, and thus more than one pattern would typically be consolidated in one consolidation period. Acquisition of the first two patterns was followed by fewer than three consolidation trials to not give these patterns too much of a head start: as there were no previous patterns for the model to choose, the first patterns were always consolidated in these early consolidation trials. Since the very first pattern was learned in an 'empty brain', it is atypical and has been excluded from all analyses and all figures showing results.

Because the model is stochastic and the number of active nodes is not always equal to the equilibrium number $k$, the model could consolidate noisy patterns, or mixtures of patterns, or no pattern at all. Nevertheless, we view the procedure followed here as more realistic than imposing the pattern to consolidate from the outside.

After learning and consolidation, the model was subjected to a test phase. Each pattern was tested a number of times by activating and clamping part of the pattern in the trace layer (the cue), and then letting the model cycle for 70 iterations. No nodes were clamped in the link layer. After the model had gone through the 70 iterations, the active trace nodes of the pattern that were not part of the cue were counted. That number was divided by the number of trace nodes in the pattern (and not part of the cue), and used as the measure of performance. Less crude measures – such as a cut-off score above which the pattern was counted as retrieved- were also tried, but did not yield qualitatively different results. After every test of a pattern, the model was reset, meaning that inhibition parameters were re-initialized and that a random set of nodes was activated. Then the cue was activated and clamped anew, and another test was carried out. The scores that will be reported are thus the average proportions of trace nodes in a pattern that were not part of the cue, and that were active after 70 iterations in the test.

**Results and discussion**

In most of the consolidation trials, one relatively intact pattern was found and consolidated. In the simulation of normal learning, this was the case in 86% of the consolidation trials. In 9% of the trials no pattern was found by the model, and in 5% more than one pattern was active at consolidation (usually two). Most often the pattern consolidated was the one learned just before the consolidation trial or the pattern before that, but it could also be an older pattern. Figure 9 shows the likelihood that a given pattern was consolidated at different times in the experiment. For all patterns, this likelihood was relatively high in the first consolidation periods after the pattern was learned, and dropped off when more and more new patterns were learned. The oldest patterns maintained an advantage, however, in the sense that their total amount of consolidation was greater than the total amount of consolidation for later patterns. This is because early in the simulation few patterns have been learned, and patterns thus have few competitors for consolidation resources. Therefore these patterns are strongly consolidated in the first trials. They keep an edge over newer patterns and are more often consolidated later in the simulation.
TRACELINK: A MODEL OF CONSOLIDATION AND AMNESIA

Figure 9: Diagram showing which patterns are consolidated in which consolidation period. The abscissa shows the pattern presented for acquisition prior to the given consolidation period. The ordinate shows the likelihood of every pattern of being consolidated in the trials in that consolidation period. These likelihoods do not always sum to exactly 100% because in some trials more than one pattern, or none at all, are consolidated (see text).

We fitted several functions on the consolidation functions found by following for the first ten patterns how much they were consolidated at each time step after learning. As is clear from Figure 9, the likelihood that a given pattern was consolidated monotonically decreases with every new pattern that was learned. A power function fitted best on the consolidation curve of some patterns, a logarithmic function on others. Fits varied from 0.91 to 0.97 for power functions and 0.92 to 0.99 for logarithmic functions.

Figure 10 shows the results of the first simulation (the filled circles). Performance was very high for the most recent pattern; the older a pattern was, the lower it scored on the test. The curve was best approximated by a power function, which explained 93% of the variance. This is in accordance with data from human subjects: the power function is often seen as the best approximation of the retention function for human memory (Anderson & Schooler, 1991; Rubin & Wenzel, 1996; Wixted & Ebbesen, 1991).

Figure 10 also shows the chance level in the simulations. Since TraceLink strives for a given number of active nodes in a layer (parameter $k$), during testing there will always be some active nodes that may or may not belong to the pattern. Usually, when the cued pattern was not activated during testing, some other pattern was. Since the patterns overlapped, it was very likely that some nodes of the target pattern were also part of any non-target active patterns. Even a non-learned pattern could thus be expected to score higher than zero because of this overlap. In fact, we determined the chance level by defining an extra pattern that was not learned, testing this pattern in the usual way, and counting how many nodes of this dummy pattern were activated by
chance. The figure shows that although forgetting was substantial, pattern recall remained well above chance.

These results show the forgetting that occurs in the model when new patterns are learned. The mechanism that accounts for forgetting in TraceLink is interference through overlap of patterns. When a pattern is learned, it is immediately stored in the link system in a way that enables retrieval later on. The cue that is presented in the trace system during the test can activate the pattern in the link system, which subsequently activates the rest of the pattern in the trace system. When a new pattern overlaps with an old one in node X, the old pattern is partly unlearned: while new connections are laid between node X and the other nodes in the new pattern, the connections from the old pattern to node X are unlearned. The node is thus effectively disconnected from the old pattern. The whole pattern is gradually unlearned when more and more nodes are disconnected from the pattern due to overlap with newer ones. Since there is more overlap in the link system than in the trace system, the link portion of a pattern is lost relatively rapidly, and the trace portion more slowly. By the time 15 patterns have been learned, many nodes in the first pattern have been part of at least one other pattern and are thus lost for the first pattern; this can be calculated to be the case for on average 54% of its trace nodes, and 94% of its link nodes. Since old patterns decay rapidly in the link system, patterns in the link system quickly lose their ability to be activated by a cue in the trace system, or to maintain a stable activation of the pattern in the trace system. During the time that the pattern is still strong in the link system, however, the pattern may be consolidated, and thus its memory strength continues to build up in the trace system. This enables retrieval of the older patterns from the trace layer. An old pattern may also be activated on the basis of a very strong trace representation alone.

![Figure 10: Results of the basic simulation of normal learning and forgetting (filled circles), and the simulation of retrograde amnesia (open squares). Fifteen patterns were learned; the patterns on the left were the most recently learned, the patterns on the right are the oldest. The first-learned pattern is not shown. Scores are the mean proportion of nodes in the trace portion of the pattern that are active at test and not part of the cue. The continuous lines are power fits of the two data series. They explain 93% of the variance in normal learning, and 92% in retrograde amnesia.](image)
Forgetting is thus explained in TraceLink by overwriting by subsequent memories. There are at least two other ways in which forgetting can be modeled: by decaying connections (e.g., Alvarez & Squire, 1994), and by changes in the set of context cues (e.g., Chappell & Humphreys, 1994; Mensink & Raaijmakers, 1988). It is likely that, if TraceLink were used to explain a broader range of memory phenomena, one or both of the other approaches would also have to be incorporated in the model.

As already stated, the forgetting function and the Ribot gradient (to be discussed later) were found not only with the parameter set used here but also with many others. Some parameter sets produced less desirable side effects, however, such as sizeable primacy effects in long term memory. This occurred because consolidation is implemented in TraceLink as a competitive process. Strong patterns have a greater likelihood of becoming active when the model cycles freely during a consolidation trial, and are thus more likely to be consolidated. Since the last pattern is the strongest in the link layer, it was usually consolidated most often (see Figure 9). However, the strength of the pattern in the trace layer also played a role in determining which pattern was consolidated. If consolidation went too fast, or if the trace layer was too large compared to the link layer, the trace layer tended to determine which pattern was consolidated. Often one pattern (usually one of the first) was then consolidated over and over again. With every consolidation it would become stronger and thus more likely to be consolidated in the next trial. This resulted in a high performance for the first patterns compared to later patterns, something that can be termed 'runaway consolidation' (also see chapter 7). We avoided this artifact by choosing the parameter set in such way that the strength of patterns in the two layers were balanced.

Simulation 2: long-term retention and permastore.

Though monotonically decreasing retention curves have been found with a broad range of time scales, there is evidence that after about 3-5 years (Bahrick, 1992), forgetting ceases and recall probability remains stationary. Bahrick has called this state 'permastore' (Bahrick, 1984, 1992; Bahrick et al., 1975). This phenomenon has been found, among others, in retention of high school Spanish (Bahrick, 1984), memory for the faces and names of classmates (Bahrick et al., 1975), and retained knowledge from a college course in cognitive psychology (Conway, Cohen, & Stanhope, 1991).

Method

To test whether the model would develop permastore, we let the model learn more patterns. Simulation 2 proceeded exactly as simulation 1, with one difference: in this simulation, the model acquired not 15 patterns but 20.

Results and discussion

Figure 11 shows the results from the permastore simulation. We fitted a power curve on the first 15 patterns (the number of patterns in the first simulation), and found that it explained 96% of the variance. If a power curve was fitted on all patterns, the fit decreased to 87% (the fit did not increase when the chance level was taken into account). On the last ten patterns, however, a flat line fitted very well. The best fitting regression line (drawn in Figure 11) had a slope of just -0.0002. The forgetting curve shown by the model thus reaches an asymptote, at which forgetting stops and performance does not further deteriorate. A balance has been reached between memory decay and memory consolidation: the gains of consolidation are balanced by the
FORGETTING caused by the acquisition of new patterns. This results in a state of permastore, where patterns remain at a constant level of retrieval performance. Perhaps permastore is thus not a state of immutable memory strength, but instead a dynamic strength in which forgetting is balanced by consolidation.

Although the comparison between raw modeling results and experimental results is inappropriate, some may note that Bahrick (1984, 1992; Bahrick et al., 1975) found that substantially more material resided in permastore after four years than is apparent in our simulations. This is not surprising, as material in Bahrick’s study was often overlearned (e.g., classmate names). With less overlearned materials, estimates of the number of memories in permastore may be much lower (Meeter, Murre et al., subm.).

5.3 RETROGRADE AMNESIA

In the TraceLink model, retrograde amnesia can be modeled by a temporary or persistent loss of link nodes. This has different effects for recent compared to older memories, as becomes clear when we contrast stage 2 memories with stage 4 memories. Stage 2 representations are dependent on a functioning link system for their retrieval and internal coherence. When the link system is disabled, the trace nodes in the pattern lack the support from the link system and cannot activate the other trace nodes in the pattern. Stage 4 memories, however, have developed a supporting trace-to-trace connectivity structure and, as was illustrated in Figure 6, their retrieval is independent of the link system. We postulate that this is the main mechanism underlying Ribot’s Law. Stage 3 memories are a case in between. Their successful retrieval after a lesion of the link system will depend on precisely which link nodes are unavailable and which trace-to-trace connections have been formed already (one could speculate that memories
within strongly associated clusters support each other's the retrieval, which may explain the occasional isolated islands of memories preserved in an otherwise dense retrograde amnesia).

Simulation 3: Retrograde amnesia and the Ribot gradient

Method
The learning phase in the simulation of retrograde amnesia proceeded as in simulation 1, with normal acquisition and consolidation of patterns. The only difference with the previous simulation resided in the test. Before the test the link layer was deactivated to simulate a lesion in the medial temporal lobe and hippocampus. This meant that the trace layer was allowed to cycle, while all nodes in the link layer were inactive.

Results and discussion
When the model was tested with a functioning link layer, performance was very high for the most recent pattern. We obtained the opposite when the model was tested with a deactivated link layer. Though performance was relatively low for all patterns, the most recent patterns suffered more from the deactivation of the link layer than the oldest patterns (see the open squares in Figure 10). As can be seen in Figure 10, the most recent pattern does not score much better than chance in the condition of retrograde amnesia. This corresponds to the Ribot gradient found in patients with retrograde amnesia.

A difference between this simulation and a typical retrograde amnesia experiment, is that in the simulations we worked with equivalent items, i.e. all the items are learned with the same strength. Our control condition shows a strong forgetting gradient. Tests of retrograde amnesia, on the other hand, are typically constructed in such a way that an equal number of items from various decades are answered correctly by normal controls (see chapter 3). Since more information from previous decades can be assumed to be forgotten, this equality of performance for normal controls implies that the items selected from the various decades differ in average initial learning strength. To compare data from the simulations with data from retrograde amnesia tests, one might plot the retrograde amnesia scores as a percentage of the normal control scores, which would make the Ribot gradient much steeper (see Figure 18 in chapter 6).

The TraceLink model is able to simulate both the forgetting function and the Ribot gradient with relatively simple assumptions. The basic mechanism that produces both curves is that performance for the most recent patterns is based on the link system, while performance on the older patterns is based on the trace system. When the link layer is intact, it enables high performance for the most recent patterns, producing the forgetting curve. When the link layer is lesioned, recent patterns that depend solely on the link layer become unretrievable. Older patterns that have been consolidated are still retrievable from the trace layer. This explains the Ribot gradient found after the lesioning of the link system.

5.4 SHRINKAGE OF RETROGRADE AMNESIA

Shrinkage refers to the process of recovery from retrograde amnesia. The term implies that in recovery from retrograde amnesia, older memories tend to become available before more recent memories (though, as mentioned in previous sections, isolated islands
may become available before certain older memories; Whitty & Zangwill, 1977). TraceLink models amnesia that later resolves by a temporary unavailability of link nodes. This might be caused by, for example, a shift in the balance of inhibition and excitation through a lesion or another abnormality, resulting in a suppression of activity in link structures. Stage 4 memories are unaffected by this unavailability, but retrieval of stage 2 and stage 3 memories is impaired. If the unavailability of link nodes persists, the only recovery possible is by self-repair of stage 3 memories that are just below the threshold of complete consolidation (see Robertson & Murre, 1999 for a general approach to self-repair and recovery of brain damage).

If link nodes do become available again, several situations are possible. It may be that synaptic weights on trace-link connections have deteriorated to low or random values. This means that stage 2 memories and stage 3 memories have been lost. The perspective for recovery is better than with a persistent link lesion, as the trace-link connections can now participate in the continuing consolidation process. If part of the original weights on the trace-link connections have been preserved, one would predict fairly rapid initial recovery (weeks) of many stage 3 memories and some stage 2 memories, followed by a long period of gradual recovery (months) of some of the remaining memories. If all trace-links become available again with little loss of weights there is no need for a time-consuming processes of self-repair of the representations. This last scenario may portray recovery from TGA, which can be very rapid (hours or less). Our simulation of shrinkage concentrates on this fourth case; we simulated a TGA attack and its resolution.

Simulation 4: Transient global amnesia (TGA)

Method
Evans et al. (1993) observed medial temporal lobe hypoperfusion during a transient global amnesia (TGA) attack. Interpreting the hypoperfusion that Evans et al. found as a sign of low activity, we simulated TGA by temporarily suppressing activity in the link layer. This was done by lowering the value of \( k \) in the link layer. The parameter \( k \) is the number of nodes that are active in equilibrium. When it is lowered in a given layer, fewer nodes will generally be active in that layer.

After 14 patterns had been learned normally, \( k \) was set to zero in the link layer to simulate the TGA attack. This means that the model will try to suppress any activity in the link layer. One pattern was then learned under this simulated TGA, without any link activity. Then the first test occurred: all patterns were tested while the \( k \) parameter was still set to zero in the link layer. To simulate the gradual lifting of TGA, \( k \) was set to 3 and the model was tested a second time. A third test occurred with \( k \) in the link layer set to 5, nearly back to its normal level of 7. To simulate completely resolved TGA, \( k \) in the link layer was set to its normal level, and the model learned five more patterns with the link layer wholly active. At the end of the simulation, the model was tested a fourth and final time, with \( k \) at its normal level.

Results and discussion
Figure 12 shows the results for the tests during simulated TGA, during its resolution, and after the attack. In the first test, the simulated TGA resulted in a dense retrograde amnesia, and in a severe anterograde amnesia for the one pattern learned during the TGA-attack (Figure 12a). There is also a Ribot gradient in the retrograde amnesia, as has been found in patients during a TGA attack ( Hodges & Ward, 1989). The suppression of
activity by the lowering of $k$ makes it impossible for patterns to reach a sustained activation in link, which has a detrimental effect mostly on the more recent patterns that depend on the link layer for their retrieval.

*Figure 12b* and *Figure 12c* show how during the resolution of the simulated TGA, more and more of the old patterns become available. The old patterns rapidly return to near-normal performance relative to the more recent ones, and even exceed their normal level (for comparison the results of simulation 1 are also drawn in the figure). We thus find the shrinkage of retrograde amnesia that is typically observed during recovery of TGA by Hodges and Ward (1989), in which the amnesia first resolves for the older memories and only later for the more recent ones. *Figure 12d* shows performance on the fourth test, after the simulated TGA-attack has resolved. Performance on all patterns is back to normal, except for the pattern that has been learned during the simulated TGA attack: performance for this pattern is extremely low. This is typical for TGA patients, who after the attack have dense amnesia for the period of the TGA itself.

The TraceLink model is able to simulate reversible amnesia and the shrinking of retrograde amnesia typically observed when the amnesia lifts. The mechanism simulated here may explain why patients after trauma often have an anterograde and retrograde amnesia that later partly resolves. Their lesions may upset the balance of activation and inhibition, and lead to an abnormally low activity level in link structures.

*Figure 12*: Performance in the simulation of Transient Global Amnesia (TGA). A grey hatched area indicates a period of TGA. (a.) Performance during the attack. The line marked “onset” indicates the onset of the attack. The model shows anterograde amnesia for the pattern learned after the onset of the TGA (pattern 1), and temporally graded retrograde amnesia for the patterns learned before the attack (patterns 2 to 14). (b.) Gradual lifting of the TGA; 25% of the link nodes have become available again. (c.) 50% of nodes are available. (d.) Performance after TGA has lifted and five new patterns have been learned (labeled 1 to 5). There is only amnesia for the pattern learned during the attack, all other amnesia has resolved.
The difficulties caused by this low level of activity (illustrated by the present simulation) may add to the memory deficits caused by the lesions. Though genuine restitution may also play a role, part of the resolution of amnesia may be caused by a return to normal activity levels in link structures. This entails the empirical claim that if patients suffer from temporary amnesia, one should be able to detect a pathologically low activity level within link structures (e.g., within structures of the medial temporal lobe).

5.5 ANTEROGRADE AMNESIA AND ITS CORRELATION WITH RETROGRADE AMNESIA

Anterograde amnesia can have two causes in TraceLink: (i) a lesion or dysfunction of the link system, and (ii) a lesion or dysfunction of the modulatory system. These two causes of anterograde amnesia have different effects on retrograde amnesia. When the link system is lesioned, the anterograde amnesia is accompanied by retrograde amnesia. When the modulatory system is lesioned, however, there is no retrograde amnesia. The fact that two lesions can lead to anterograde amnesia, and have different consequences for retrograde amnesia, makes it possible for TraceLink to explain the correlations between anterograde and retrograde amnesia observed in different populations.

Simulation 5: Anterograde amnesia through lesioning the link layer

Method
The first cause of anterograde amnesia in TraceLink is a lesion to the link system. This can be simulated by deactivating the link layer, followed by learning patterns with only the trace layer functioning. We simulated four degrees of lesioning of the link layer: of the entire link layer, three quarters of the link layer, one half of the link layer, and only one quarter of the link layer. In all simulations, the model learned 12 patterns before the lesion occurred. Three more patterns were learned while the link layer was lesioned, after which the model was tested.

Results and discussion
Figure 13 shows the effects of lesioning the link layer. For comparison, the line of normal forgetting (see simulation 1) is also drawn. A lesion of the whole link layer (Figure 13a) caused an extensive retrograde amnesia with a Ribot gradient, and also a near complete anterograde amnesia. Without a functioning link system, recently learned, not yet consolidated patterns are lost, and new patterns cannot form stable representation. These deficits are attenuated if some link nodes are still available. Lesioning three quarters of the link nodes (Figure 13b) produced a serious anterograde amnesia, and retrograde amnesia for a sizeable number of patterns preceding the lesion but without a clear Ribot gradient. Lesions of one half of the link layer (Figure 13c) caused a mild to moderate anterograde amnesia, and just a brief, mild, retrograde amnesia. A lesion of one quarter of the link layer (Figure 13d) produced almost no retrograde amnesia, and only a mild anterograde amnesia. TraceLink thus predicts a correlation of anterograde amnesia and retrograde amnesia when the cause of the anterograde amnesia is a lesion of the link system. Also, Ribot gradients only appear in TraceLink when there is a near-total lesion of the link system. If the lesion is less than complete, no absolute Ribot gradient appears, in the sense that recent memories are worse than equivalent remote memories. This can
be seen as a prediction for the animal literature, in which equivalent memories can be used. In patient studies, however, retrograde amnesia tests are used in which items are of equivalent recall probability. Accounting for forgetting, this means that remote memories must actually have been stronger at encoding. As described above, patient data can therefore best be compared to the results of the lesioned model divided by the results of the control simulations. As inspection of Figure 13 will show, TraceLink therefore still predicts a notable relative Ribot gradient in scores on typical retrograde amnesia tests if the Link lesions are sizeable but not complete (75%, 50%).

Moreover, our model predicts that small link lesions produce mild anterograde amnesia accompanied by hardly noticeable retrograde amnesia. This may explain the deficits of a category of patients that is usually referred to as ‘showing isolated anterograde amnesia’. Patients with lesions limited to field CA1 of the hippocampus tend to show mild to moderate anterograde amnesia, and only mild retrograde amnesia for a limited period before the lesion (Rempel-Clower et al., 1996; Zola-Morgan et al., 1986). Such a pattern is shown by the model when only one fourth of the link layer is lesioned.

![Figure 13: Results of the simulation of anterograde amnesia through lesioning of the link layer. In all simulations, the model learns 12 patterns, labeled 4 to 15, before the lesion occurs. After the lesion (marked by the line labeled “onset”), the model learns 3 more patterns. (a). Performance after deactivation of all of the link layer; (b). deactivation of 75% of the link layer; (c). of 50% of the link layer; (d). of 25% of the link layer.](image-url)
Simulation 6: Anterograde amnesia through lesioning the modulatory system

Method
Another way that anterograde amnesia can occur in the TraceLink model is by a lesion of the modulatory system. Since the modulatory system is not implemented in the simulations, we simulated the effects of such a lesion by a decrease in the plasticity of the link system. The learning rate in the link layer was, after the lesion, set at a base rate value equal to the learning rate in the trace layer (see Table 5). The simulation started with the normal acquisition and consolidation of 12 patterns. Then the simulated lesion of the modulatory system was applied. Three more patterns were presented to the model after the lesion, after which the model was tested.

Because the modulatory system in TraceLink has a role in regulating plasticity during consolidation, it is unclear whether after a lesion of this system consolidation continues in patients. We therefore simulated two conditions. In the first condition, consolidation continued after the lesioning of the modulatory system; between acquisition of two patterns, there was a normal consolidation phase. In the second condition, there was no more consolidation after the lesion.

Results and discussion

Figure 14 shows the effect of lesioning the modulatory system. For both the condition with and without consolidation after the lesion, lesioning the modulatory system produces very dense anterograde amnesia, but no retrograde amnesia at all. On the contrary, performance for patterns learned before the lesion is even better than in the simulation of normal forgetting (which is drawn in the figure for comparison). This has a simple explanation. In the normal simulation, the model learns new patterns that interfere with the older ones, which induces forgetting and a lower performance for the older patterns. When the modulatory system is lesioned, however, new patterns are learned only faintly in the link layer, so that these patterns interfere little with the patterns that are already engrained in the link layer. Therefore little forgetting occurs in the link layer for the patterns learned before the lesion, and these patterns remain strong.

Comparison of the two conditions, with and without consolidation after the lesion, shows that the afore-mentioned effect is weaker when there is consolidation after the lesion has occurred. This might be caused by the learning during consolidation, which may have an unlearning effect in some cases. Specifically, when of a pair of nodes just one is active, the connection from the inactive node to the active one is weakened (see the Appendix to this chapter). Nevertheless, in both conditions performance for the patterns learned before the lesion is higher than in the normal simulation.

TraceLink thus makes the rather counterintuitive prediction that there is a class of patients that have dense anterograde amnesia and no retrograde amnesia, but instead have better-than-normal memory for the events in the time right before the lesion. These would be patients in which the structures that correspond with the modulatory system are lesioned, but none that correspond to the link system. Up to this moment, no such patient has been reported in the literature: every patient with enduring moderate to severe anterograde amnesia has at least a short period — sometimes just a few weeks— of retrograde amnesia (see the review of amnesia). Possibly, lesions limited to the modulatory system do not occur because link structures and modulatory system structures are overlapping. However, the predictions given here could be tested using scopolamine, a drug that produces anterograde amnesia without retrograde amnesia.
Figure 14: Results of the simulation of anterograde amnesia through lesioning of the modulatory system. After 12 patterns have been learned, the modulatory system was lesioned. “Control” refers to the control simulation without anterograde amnesia (simulation 1). “Consolidation” refers to the simulation in which the model continued consolidating patterns after the lesion, “No consolidation” to the simulation in which that did not occur. After the lesion (marked by the line labeled “onset”), the model learned 3 more patterns.

(though scopolamine would have to be administered for a longer period than it typically is).

The correlation between retrograde and anterograde amnesia

Two mechanisms can lead to anterograde amnesia in the TraceLink model: a lesion or dysfunction of the link system, and a lesion or dysfunction of the modulatory system. Together, these mechanisms can explain the whole spectrum of anterograde-retrograde correlations. Loss in the link system, on the one hand, will cause both retrograde and anterograde amnesia, and their severity will show a correlation: as more link nodes and trace-link connections are lost, more existing memories become unretrievable, and it will be harder to form new memories. Dysfunction of the modulatory system, on the other hand, makes it hard to form new representations, but has no effect on the existing ones.

For lesions limited to structures corresponding to the link system (such as the parahippocampal region), TraceLink predicts a strong correlation between anterograde amnesia and retrograde amnesia, and also between the latter two and the size of the lesion. Diffuse lesions that affect the link, trace and modulatory systems equally will show more variability in the pattern of amnesia, but in general anterograde and retrograde amnesia would still be expected to correlate with each-other, and with the severity of the lesion. This may explain the correlations found with closed-head injury.

If effects of the lesion are disproportionally large in the modulatory system, we would predict a disproportional anterograde amnesia relative to retrograde amnesia. We believe that this is the case for Alzheimer’s patients, in whom disproportionate
anterograde amnesia is often a feature of the first stages (Spaan, in prep.). With Alzheimer's disease, the hippocampus and basal forebrain usually receive disproportional lesions, relative to other medial temporal lobe structures (Hyman, Hoesen, Damasio, & Barnes, 1984; Van Hoesen, 1990; Whitehouse et al., 1982). The former are likely candidates for the modulatory system, whereas the latter—with the hippocampus—may be involved in the link system. This means that the correlation between anterograde amnesia and retrograde amnesia, caused by damage to the link system (medial temporal lobes, including the hippocampus), is upset by additional, unrelated damage to the modulatory system (hippocampus, basal forebrain). In Alzheimer's disease, we therefore expect much lower correlations between anterograde and retrograde amnesia than, say, in the case of closed-head injury.

5.6 ISOLATED RETROGRADE AMNESIA

Isolated retrograde amnesia, also known as focal retrograde amnesia (Kapur, 1993), is retrograde amnesia without accompanying anterograde amnesia. It can be simulated in TraceLink by a rupture or deterioration of the connections between the link system and the trace system. The weights (strengths) of these connections code for the learned patterns, and when they are lost, retrograde amnesia occurs. If, after such loss, it is still possible for the system to form new connections between the trace and link systems, new memories can still be formed and little anterograde amnesia need occur. It is unlikely, however, that immediately after a rupture of old connections new connections can be formed. We assume that it takes some time before enough synapses are available (Robertson & Murre, 1999). This can explain the initial anterograde amnesia that later resolves, leaving isolated retrograde amnesia.

Simulation 7: Lesioning the connections between trace and link

Method

The model initially learned 12 patterns, after which the connections between the link layer and the trace layer were severed. To simulate a partial lesion, all weights of these connections were multiplied by a random factor varying with a uniform distribution function between 0 and 0.20. After the lesion, weights were thus only between 0 and 20% of their initial value (this way of modeling a connectivity lesion would be appropriate if, for example, each connection consisted of a number of axons of which a certain percentage would be cut). After this, the model learned four more patterns. To simulate the gradual reappearance of connections, the learning rate for the connections from trace to link and vice versa were first set to half its usual value, and moved back to its usual value with an exponential function \((1-0.5^x)\), where \(x\) is the number of the pattern after the lesion). Finally, the model was tested.

Results and discussion

Figure 15 gives the result of the simulation and as a comparison the control simulation. As can be seen, the lesion resulted in a dramatic loss of pre-lesion patterns. The first patterns learned after the lesion were not learned well due to the lower learning rate in the connections between the trace- and link layer. This corresponds to a retrograde amnesia and an anterograde amnesia immediately after the lesion. For the most recent patterns, performance was normal. The model did not suffer from any residual
anterograde amnesia. The retrograde amnesia therefore corresponds to an isolated retrograde amnesia.

Because cases of isolated retrograde amnesia are rare, there is currently little knowledge about the precise neurological basis of the syndrome. It is thus not yet possible to validate our approach toward isolated retrograde amnesia against neurological data. The structures implicated in isolated retrograde amnesia, anterior structures of the temporal lobe such as the parahippocampal region and the temporal pole, are structures that link the hippocampus with areas in the neocortex. Squire and Alvarez (1995) favor the theory that with isolated retrograde amnesia the knowledge base itself is damaged, leading to dense retrograde amnesia, with flat temporal gradients. There is good evidence that this pattern occurs in some patients (Kapur, 1993). In other cases, however, the isolated retrograde amnesia shows evidence of a Ribot gradient (see the review of retrograde amnesia in Section 2). A combination of Squire's hypothesis and the one offered above may be needed to explain all cases of isolated retrograde amnesia.

5.7 IMPLICIT MEMORY IN AMNESIA

When a new pattern is learned, strong connections are formed between activated nodes in the trace system and the link system, and among the activated nodes in the link system. Some learning also occurs between the nodes in the trace system, though at a much lower rate than for the connections within link or between trace and link. This learning is, on itself, not enough to sustain a new memory after a single trial, as was shown in the simulations of anterograde amnesia.

![Figure 15: Results of the simulation of isolated retrograde amnesia. After 12 patterns had been learned, connections between the trace layer and the link layer were disturbed. This is marked by the line labeled "onset". After the lesion, patterns were learned with a lower learning parameter for the connections between trace and link. The learning parameter moved back to its normal value with an exponential function (see text for details).](image-url)
In many cases, however, trace nodes will be activated that already have connections between them. In these cases, the already existing connections will be strengthened. This strengthening of existing memories in the trace system, as opposed to forming new memories via the link system, is how TraceLink models implicit learning. The increment of the connection strengths in the trace system may also offer a way to model the acquisition of new skills and of low-level knowledge in amnesic patients. Every presentation of a pattern leads to a small increment in the strength of connections in the trace system. Slowly these increments may accrue until a strong bond is formed between the nodes in a pattern.

The link system and modulatory system are both crucial for the formation of new connections, but need not be involved in further strengthening of existing connections. If implicit learning were dependent on the modulatory system, we would expect implicit memory tasks to be sensitive to modulating factors such as arousal, which they are not (Gold, 1995; Jacoby & Dallas, 1981). The link system and modulatory system can nevertheless be active during this process, but this merely implies that implicit and explicit memories typically form together in normal subjects (Jacoby, 1991). Indeed, with a few exceptions, whatever representations are being strengthened by implicit memory processes, these have been formed as an explicit memory process at some earlier time. With a failing link system or modulatory system, implicit learning is still intact because strengthening of existing connections is independent of the link system and the modulatory system. Implicit memory will, therefore, be preserved in amnesia.

Simulation 8: Implicit memory

Method

Our simulation of implicit learning was kept relatively simple. We simulated implicit learning as learning without involvement of either the link layer or the modulatory system. First, 15 patterns were learned with normal involvement of all systems in the model, and with normal consolidation. This simulated the learning of a subject who comes into an experiment. After all patterns had been acquired, two randomly selected patterns were given an extra, simulated implicit learning trial. In this trial, only the trace portion of the pattern was activated. Learning occurred within the trace layer, with the normal trace learning parameter (see table 2). After the two patterns had received an implicit trial in this way, the model was tested twice: first with the link layer functioning to simulate implicit memory in normal controls, then with the link layer deactivated-to simulate implicit memory in amnesic patients. The simulation was replicated 1350 times to generate enough data points for every pattern, since in one replication just two patterns received an implicit learning trial.

Results and discussion

Figure 16 shows the results of the simulation, contrasting performance before an item had received an implicit learning trial, with performance after an implicit learning trial. As a comparison, the learning of a new pattern via an implicit learning trial is also shown. The implicit learning trial had a substantially greater impact for the already learned patterns than for the new pattern. New patterns benefited more than old patterns, though the effect was small. Furthermore, the impact of the implicit learning trial was greater in the condition with a deactivated link layer than in the normal condition. This can easily be understood, as in the normal condition performance is
partly a function of the (unchanging) strength of the pattern in the link layer. The most important conclusion, however, is that implicit learning trials as implemented in the TraceLink model benefits both artificial normal controls (results of the normal test) and artificial retrograde amnesia patients (results of the retrograde amnesia test).

5.8 VARIATIONS IN ENCODING

In all simulations already discussed, the model learned equivalent, equally strong patterns. Variations in strength emerged during the simulation because some patterns had more overlap with the rest of the patterns than others (all patterns were random) or because some patterns were chosen more often in the random consolidation process, and not because of differences in the original learning conditions. It is, however, not realistic to assume that all patterns are learned at acquisition with equal strength. We therefore investigated variations in initial pattern strength, which can be seen as reflecting the effects of arousal on memory.

Simulation 9: The effect of arousal

In TraceLink, variations in the strength of acquisition are caused by the modulatory system, the system that modulates learning in the link system. High activation of this system results in a higher learning rate in the link system, and thus stronger connections between the nodes in the link part of a pattern. In the present simulation, activation of the modulatory system did not change the learning rate in the trace system. This choice was motivated by parsimony, but there is also empirical evidence for a lesser modulation of learning in the neocortex compared to learning in the hippocampus. Neuromodulators that enhance explicit memory, for example, have no detectable influence on performance in implicit memory tasks (Gold, 1995).

Figure 16: Results of the simulation of implicit memory. Plotted is the enhancement in performance after an implicit learning trial (priming effect), on top of base level performance. For the Control-new pattern refers to a new unlearned pattern receiving an implicit learning trial. Base level is here random activity of the pattern nodes.
Method

The modulatory system was not implemented in this simulation, but we simulated its functions by manually setting the learning rate in the link layer to a higher value for some patterns, and to a lower one for others. The learning rate in the link layer, induced by the modulatory system, varied with a continuous distribution function around the generic learning rate for the link layer given in Table 5. In one simulation, it varied between 90% and 110% of the value given in Table 5 (the low variance simulation), in another it varied between 50% and 150% of that value (the high variance simulation).

For reasons of computational economy, 18 patterns were learned in a simulation instead of the usual 15. The model was tested twice, once with a functioning link layer, and once with a deactivated link layer to investigate the effect of variance in pattern strength on the Ribot gradient.

Results and discussion

To see what influence the variation in strength had, we did an analysis of variance on the results of the simulation. We calculated the percentage of the variance in the results of the normal test that was explained by two variables: first, the variations in the learning rate, and second, the order in which the patterns are learned, i.e. whether the pattern was old or a recent at the moment of the test (see Table 6). When all patterns are learned with equal strength, a large part of the variance in the simulations is explained by whether a pattern is learned recently or some time ago (and thus already partially unlearned). When patterns are not all equally strong, however, much of the variance is explained by the variation in initial strength, and only little by the order in which the patterns were learned.

This translates into a much flatter forgetting curve (see Figure 17a, open triangles): strongly learned old patterns do not necessarily perform worse than more recent but weaker patterns. The Ribot gradient, shown in Figure 17b, is not flatter but instead slightly steeper when variance is introduced in the strength of patterns. We divided the

<table>
<thead>
<tr>
<th>variance explained by...</th>
<th>no variance strength</th>
<th>small variance strength</th>
<th>strong variance strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>age' of a pattern</td>
<td>38%</td>
<td>26%</td>
<td>6.50%</td>
</tr>
<tr>
<td>strength of a pattern</td>
<td></td>
<td>8%</td>
<td>33%</td>
</tr>
<tr>
<td>interaction btw. 'age' &amp; strength</td>
<td>*</td>
<td>5%</td>
<td>6.50%</td>
</tr>
<tr>
<td>error</td>
<td>62%</td>
<td>61%</td>
<td>54%</td>
</tr>
</tbody>
</table>
Figure 17: Results of the simulation of the effects of arousal on memory. (a.). Contrast between results with strong variance in learning rate and with no variance in the learning rate. (b.) Same as panel a. for the lesioned model. (c). Simulation with strong variance in learning rate, with patterns subdivided into those learned either with a high, low or medium learning rate (Strong=learned with 150% to 120% of the normal learning rate, Middle=80% to 120%, Weak = 50% to 80%). (d). Same as panel c. for the lesioned model.

patterns in the high variance simulation into strong patterns, learned with an initial strength between 120% and 150% of the normal value, medium patterns, learned with between 80-120% of the normal learning rate, and weak patterns with strength between 50% and 80% of the normal learning rate (see Figure 17c and Figure 17d). This showed that the strong patterns were not only better recalled in the immediately after learning, but that the effect persists for old patterns. For the normal simulation, high initial strength of patterns results in a long-term primacy effect, in which the oldest patterns have an advantage over less remote ones (see Figure 17c). This effect is reminiscent of the primacy effect sometimes observed in remote memory. Sehuster (1989) examined his memory for opera performances in his 25 years as a season ticket holder, and found a typical serial position curve with recency and primacy effects: his memory was best for the most recent ones, but it was also good for the first seasons in which he held season tickets. For the Ribot curve, in the high variance simulation it can be attributed almost completely to the strong patterns (Figure 17d, filled triangles). There was no Ribot gradient in the performance of the weak patterns.
The results thus show that the variations in strength had a prolonged influence on the retrievability of the pattern. This is surprising, since only the initial strength of the link portion of the patterns was varied, and this portion is unlearned relatively fast. The explanation for this finding is that strong patterns tend to monopolize consolidation resources. If weak patterns are learned after a strong one, the strong one tends to be the one that is consolidated. Though the strong pattern is subsequently unlearned in the link system, its trace portion tends to become so strong that the pattern remains a likely winner in the competition for consolidation resources. We checked which patterns were consolidated in the simulation with high variance, and found that 85% of the consolidated patterns were strong patterns, 14% medium patterns, and only 1% consisted of weak patterns. Through consolidation, the patterns with strong connections in the link system thus also become strong in the trace system.

TraceLink thus predicts that stronger patterns, e.g. emotional ones, are consolidated more often, and thus are forgotten more slowly than weaker, e.g. non-emotional, patterns. Slower forgetting for episodes learned under high arousal has indeed been observed (Burke, Heuer, & Reisberg, 1992; LaBar & Phelps, 1998; Phaf, 1994). The finding that strong patterns remain strong through more-than-average consolidation may hint at the process behind flashbulb memories, memories of gripping events that years later seem to be remembered with as much detail and clarity as on the first day (Brown & Kulik, 1977). If consolidation indeed works as a process in which patterns compete for resources, then TraceLink predicts that strong patterns will win this Darwinian competition and that the difference between strong and weak representations will grow with time.

5.9 GENERAL DISCUSSION

We have presented in this chapter a model of normal memory and amnesia. The model is able to simulate the normal forgetting curve, and several forms of amnesia. It does so on the basis of a few assumptions that follow logically from the three dilemmas that were presented in the first chapter: (1) The forgetting paradox, the paradox that more recent memories are the easiest to retrieve but the most vulnerable to brain damage, led to the assumption of a consolidation process. (2) The connectivity problem, which refers to the fact that the encoding of episodic memories requires long-term cortico-cortical connections that can only be found sparsely, was tackled by adopting a hierarchical structure in the model, with a link system that helps retrieve patterns in the trace system. (3) The stability-plasticity dilemma inspired a modulatory system that influences link plasticity, making it higher if the circumstances suggest advantages to rapid new storage.

What TraceLink shares with previous work

These assumptions are not highly original, nor do they have to be: TraceLink aims to encapsulate and implement in some detail the main lines of thought about memory, amnesia and the brain that can be traced in the neuropsychological literature of the past fifty years. Since Murre started working on the TraceLink model (Murre, 1994), several models of amnesia have been published, most of which focus on the role of the hippocampus (e.g., Alvarez & Squire, 1994; McClelland, McNaughton, & O'Reilly, 1995). The TraceLink model itself is inspired by earlier theoretical and modeling work (e.g., Marr, 1971; Milner, 1989; Rolls, 1990; Squire, 1992; Squire et al., 1984; Wickelgren,
TRACELINK: A MODEL OF CONSOLIDATION AND AMNESIA

1974, 1979). Moreover, its overall architecture is similar in aspects to that proposed by many other researchers, in particular its broad hierarchical structure (e.g., Brown & Zador, 1990; Eichenbaum et al., 1992).

The TraceLink model is similar to one by Alvarez and Squire (1994). The latter model also uses a hierarchical structure, with a link system connecting two cortical modules, and whereby a gradual consolidation process causes a movement from 'cortico-hippocampal' dependence to a purely 'cortico-cortical' basis. This type of consolidation process is similar to one we have adapted. A difference is that the consolidation is not completely autonomous in the model by Alvarez and Squire (1994): a random hippocampal representation is activated that then activates the neocortical memory to consolidate. Moreover, the model of Alvarez and Squire (1994) is very small: just two, nonoverlapping memories are stored in the system.

McClelland et al. (1995) emphasize another rationale for two memory systems. They point out that purely sequential learning may not lead to useful internal representations, and that a case can be made for the necessity of a more interleaved mode of learning. In particular, newly learned deviant patterns may disturb already learned representations. Occasional learning trials for old patterns counteract this disturbance. There are thus good reasons for a slow (interleaved) learning process such as slow consolidation. Unfortunately, the examples by McClelland et al. (1995) are based on backpropagation models, which are very sensitive to interference by new learning, the so-called catastrophic interference (McCloskey & Cohen, 1989; Ratcliff, 1990). Several neural network models have been developed that do not suffer from this, for example ART networks (Carpenter & Grossberg, 1988; Grossberg, 1976, 1987b) and CALM modules (Murre, 1992). In addition, many improvements of backpropagation have been proposed that make it much less sensitive to interference (French, 1992; Kruschke, 1992).

What is different about TraceLink

The consolidation assumption has been defended or used by many theoreticians. This prompted Nadel and Moscovitch (1997) to call consolidation theory the 'standard model' of retrograde amnesia. TraceLink comes close to the core of that standard model. However, several features make TraceLink stand out among other incarnations of it. TraceLink is the first model of amnesia in which consolidation has been implemented as an automatic process with random cueing of remote memories. If consolidation indeed takes place during rest or during certain sleep stages, then consolidation must occur without outside steering. In the process, we discovered that the stability of such a system may easily be compromised, giving rise to 'runaway consolidation' (see chapter 7).

Another novel element of TraceLink is the inclusion of a modulatory system, proposed as a remedy to the stability-plasticity problem. Inclusion of such a system allows for a better explanation for patterns of correlation between, anterograde and retrograde amnesia. Without it, the two forms of amnesia can only correlate perfectly. Moreover, it allowed for the exploration of the effect of arousal on consolidation, and the discovery of the strong interaction between memory strength and consolidation that may explain the time course of the effect of emotional content on memories.

The sheer breadth of phenomena that TraceLink has been applied to also makes it different from other models. TraceLink shows how the 'standard view' can account for anterograde amnesia, correlational patterns, shrinkage, intact implicit memory, and other neuropsychologically relevant characteristics, while remaining consistent with characteristics of normal forgetting.
The connectionist simulations themselves were kept on a qualitative level: we did not attempt to quantitatively fit concrete data sets. This is mainly because the data in the neuropsychological literature are generally noisy due to the low numbers of patients available for testing. This problem is further aggravated by an understandable tendency among neuropsychologists to focus on interesting single cases. Instead of fitting concrete data sets, we therefore tried to simulate the direction and shape of some of the principal effects reported in the literature. We are currently developing a mathematical theory of learning, memory and amnesia that may be more amenable to quantitative fits of noisy and distorted data sets (Murre et al., in prep.).

APPENDIX TO CHAPTER 5

The model is based on binary, stochastic nodes that fire synchronously. The firing thresholds of the nodes in a module are controlled by a threshold control mechanism: inhibition in a module is diminished if there are not enough activated nodes (i.e., less than some target number \( k \)) and increased if there are too many. At each iteration, after all node activations have been updated, a learning rule is applied to all connections. The details of these mechanisms are described below. All simulations were done with the Nutshell simulator (available free of charge via http://nutshell.neuromod.org).

Activation rule

A node \( i \) has an activation \( a_i \) that can take on either of two values: 0 or 1. The probability that node \( i \) will 'fire' (i.e., that its activation becomes 1) increases with its net input, as follows:

\[
P_i = \frac{1}{1 + e^{-\text{temp} \cdot \text{net}_i}}
\]

(1)

where \( \text{net}_i \) is the total input activation to node \( i \), or the weighted input to node \( i \) minus inhibition:

\[
\text{net}_i = \sum_{j=1}^{n} w_{ij} a_j - \text{inhibition}
\]

(2)

where \( w_{ij} \) is the connection weight from node \( j \) to node \( i \), \( a_i \) is the activation value of node \( j \), and \( n \) is the number of nodes in the model (if there is no connection between \( j \) and \( i \), \( w_{ij} \) is zero by default). Inhibition is discussed in the next paragraph. The temperature parameter \( \text{temp} \) in Equation 1 controls the degree of randomness of the nodes: if \( \text{temp} \) is near zero the nodes behave as simple threshold devices, if \( \text{temp} \) is high the role of the net input is limited and the node takes on values 0 or 1 randomly. We used a temperature of 0.2 in all simulations.
Threshold control

The total number of activated nodes in a module (called \( A \)) is constantly monitored and firing thresholds are adjusted to ensure that this number does not wander too far from the target number \( k \). Each module has its own \( k \), and inhibition control in a module is independent of that in other modules. To keep the number of activated nodes \( A \) as close as possible to the target number \( k \), two thresholds \( T \) and \( \tau \) are constantly adapted. Inhibition is the sum of a fast changing threshold parameter \( T \) multiplied by the number of active nodes \( A \), and a slow moving threshold \( \tau \):

\[
inhibition = TA + \tau \tag{3}
\]

\( T \) may reflect the excitability of the basket cells by the excitatory neurons. Slow inhibition, modeled by the threshold \( \tau \), may reflect the autonomous activity of inhibitory cells. The control of fast inhibition, \( T \), is straightforward: if the total activation at time \( t \) \((A^t)\) is higher than \( k \), \( T \) is increased (more inhibition); if \( A^t \) is lower it is decreased:

\[
\begin{align*}
\text{if } A^t &> (1+\text{crit})k \\
T &= T + \Delta t \\
\text{if } A^t &< (1-\text{crit})k \\
T &= T - \Delta t
\end{align*} \tag{4}
\]

where \text{crit} is the criterion for deciding whether \( A^t \) is much larger or smaller, and \( \Delta t \) is the change made to \( T \) \((\text{crit} = 0.20, \text{and } \Delta t = 0.01)\). If \( A \) is only a little bit larger or smaller than \( k \) \((\text{e.g.}, \, k < A^t < (1+\text{crit}) \times k)\), then only \( 1/3 \)rd of \( \Delta t \) was added to or subtracted from \( T \).

One disadvantage of this method is that \( T \) may change too quickly, causing violent oscillations in activity. To prevent this, \( A^t \) is dampened by making it a moving average of the current activation and the activation of previous iterations. When \( A^* \) is the current level of activation, the value used to compute both the level of inhibition \( A_T \) and the change in the parameter \( T \) is:

\[
A^t = 0.5A^t, + 0.5A^*, \tag{5}
\]

This precedes calculation of the new threshold \( T \) (Eq. 4).

The slow inhibition process aims to keep the 'slow threshold' \( \tau \) equal to \( TA \). When the equilibrium is disturbed, for example, if the activation is diminished due to a lesion, \( \tau \) slowly decreases to a new equilibrium value. The speed of this change is determined by the parameter \( \Delta \tau \), which is chosen low \((0.001)\). The expression for calculating \( \tau^t \) at \( t+1 \) is:

\[
\tau^{t+1} = (1-\Delta \tau)\tau^t + \Delta \tau TA \tag{6}
\]

The amount of 'fast' inhibition is bounded by a minimum value \( T^{\min} \) and a maximum value \( T^{\max} \). If \( T < T^{\min} \) it is set to \( T^{\min} \), and if \( T > T^{\max} \) it is set to \( T^{\max} \). Similarly, \( \tau \) is also kept between upper and lower bounds: if \( \tau < \tau^{\min} \), \( \tau \) is \( \tau^{\min} \); if \( \tau > \tau^{\max} \), \( \tau \) is \( \tau^{\max} \). \( T^{\min} \) and
\( \tau^{\min} \) were set to 0. \( \tau^{\max} \) and \( \tau^{\max} \) were set to such high values that they were never reached in the simulations.

Learning rule
The learning rule is a simple Hebbian rule that also allows decreases in weight. The change in weight \( \Delta w_{ij} \) on each time step is equal to:

\[
\Delta w_{ij} = \mu^a a_i a_j - \mu^* a_i (1 - a_j).
\]

where \( \mu \) and \( \mu^* \) represent the learning rates. Both \( \mu \) and \( \mu^* \) must be larger than 0. The weights \( w_{ij} \) are kept within the interval \([0,1]\) by setting \( w_{ij} = 1 \) if \( w_{ij} > 1 \), and \( w_{ij} = 0 \) if \( w_{ij} < 0 \).

Parameter settings
The values of what one could call ‘equilibrium parameters’, such as the temperature, are given in the text above. The function of these parameters is to keep the number of active nodes as close as possible to the equilibrium value \( k \), and to prevent wild swings in the number of active nodes from one iteration to the next. The remaining parameters, those that may influence results more directly than equilibrium parameters, are listed in Table 5.

The learning rates are given per iteration. During learning, each pattern was learned for one iteration. To limit the impact of random fluctuations in activity level on the consolidation process, consolidation is stretched out over 8 iterations at a low learning rate (see Table 5). Total learning in the trace system during one consolidation period was the same as that during one acquisition period.