



Doubly Modifiable Synapses: A Model of Short and Long Term Auto-Associative Memory

A. R. Gardner-Medwin

Proceedings of the Royal Society of London. Series B, Biological Sciences, Volume 238, Issue 1291 (Nov. 22, 1989), 137-154.

Your use of the JSTOR database indicates your acceptance of JSTOR's Terms and Conditions of Use. A copy of JSTOR's Terms and Conditions of Use is available at <http://www.jstor.ac.uk/about/terms.html>, by contacting JSTOR at jstor-info@umich.edu, or by calling JSTOR at (888)388-3574, (734)998-9101 or (FAX) (734)998-9113. No part of a JSTOR transmission may be copied, downloaded, stored, further transmitted, transferred, distributed, altered, or otherwise used, in any form or by any means, except: (1) one stored electronic and one paper copy of any article solely for your personal, non-commercial use, or (2) with prior written permission of JSTOR and the publisher of the article or other text.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

Proceedings of the Royal Society of London. Series B, Biological Sciences is published by The Royal Society. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.ac.uk/journals/rsll.html>.

Proceedings of the Royal Society of London. Series B, Biological Sciences
©1989 The Royal Society

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2001 JSTOR

Doubly modifiable synapses: a model of short and long term auto-associative memory

BY A. R. GARDNER-MEDWIN

Department of Physiology, University College London, London WC1E 6BT, U.K.

*(Communicated by H. B. Barlow, F.R.S. – Received 9 March 1989 –
Revised 31 May 1989)*

Synapses that can be strengthened in temporary and persistent manners by two separate mechanisms are shown to have powerful advantages in neural networks that perform auto-associative recall and recognition. A multiplicative relation between the two weights allows the same set of connections to be used in a closely interactive way for short-term and long-term memory. Algorithms and simulations are described for the storage, consolidation and recall of patterns that have been presented only once to a network. With double modifiability, the short-term performance is dramatically improved, becoming almost independent of the amount of long-term experience. The high quality of short-term recall allows consolidation to take place, with benefits from the selection and optimization of long term engrams to take account of relations between stored patterns. Long-term capacity is greater than short-term capacity, with little or no deficit compared with that obtained with singly modifiable synapses. Long-term recall requires special, simply implemented, procedures for increasing the temporary weights of the synapses being used to initiate recall. A consolidation algorithm is described for improving long-term recall when there is overlap between patterns. Confusional errors are reduced by strengthening the associations between non-overlapping elements in the patterns, in a two-stage process that has several of the characteristics of sleep.

1. INTRODUCTION

Modifiable synaptic connections can, in appropriate models, give good memory performance in associative and pattern classification tasks. It is shown here that there are advantages in auto-associative tasks if connections can be modified in two independent ways, with their strengths given by the product of two parameters, termed ‘temporary’ (T) and ‘persistent’ (P) weights. The modifications occur with different time courses and conditions, corresponding broadly to ‘short term’ and ‘long term’ memory in a network.

Sections 1 and 2 in the paper are largely non-mathematical and outline principles substantiated by simulation in §§3 and 4.

(a) Temporary and persistent memory

There are three basic advantages that can arise in memory models with temporary and persistent changes. These advantages are central to the concept of a

'working memory' (Baddeley 1986) in which sensory, scratch-pad and long-term memory stores are integrated into a working whole.

Firstly, transient memory allows recent events to be stored and forgotten without saturating the long-term memory. Secondly, it permits conditions after learning to determine whether or not a long-term memory is established, thereby reducing the required capacity. Thirdly, it allows the form of long-term storage to depend on processing after the initial learning event, for example, at higher 'levels of processing' (Craik & Lockhart 1972).

An additional, more subtle, benefit from temporary changes was shown by Hinton & Plaut (1987). This can operate when subsequent learning overlays and compromises the effects of prior learning. By using connection weights that were the sum of fast and slowly adjusting components, Hinton & Plaut showed that retraining on a few patterns could tune performance temporarily to different epochs in the net's experience. This 'deblurring' effect generalized to patterns that were learned in blocks along with the retrained patterns. The effect derives from a property of error correction algorithms, whereby the engram for one pattern depends on the other patterns repeated within the same training block. Other algorithms, for example, associative algorithms, are not sequence dependent in this way and may not show such effects.

(b) *Mechanisms for storage*

This paper is concerned with learning and recall for patterns that are presented once. The initial storage algorithm must operate with a single presentation. Auto-associative nets offer a simple way of achieving this, within constraints that are fairly well understood through analysis (Willshaw *et al.* 1969; Gardner-Medwin 1976; Palm 1988) and simulation (Kohonen 1984; Lansner & Ekeberg 1985).

Once a pattern is stored temporarily, different algorithms become available for consolidation of long-term memory. These may involve repetitions of internally generated versions of the original pattern, and may produce changes either in the same structure or elsewhere. Doubly modifiable synapses permit storage of both temporary and persistent engrams in the same structure, as associations between the same elements.

Persistent storage with a different representation may sometimes have benefits, for example greater economy or less interference. Thus a scene or a sentence may initially be recalled in sensory terms and later in terms of significance. Both types of memory could plausibly involve auto-association at different levels. However, conversion to a new logical representation requires the application of rules that are not always available: we may fail to interpret the scene or understand the sentence. If conversion is impossible or potentially unreliable, then storage may best remain in its current form as an association of individual parts of a pattern. We often try to recall things in this manner, free from possibly spurious interpretations: most notably in the witness box. This paper aims to identify benefits and constraints that apply to simultaneous short- and long-term associative memory at a single level of representation.

(c) *Generalization and the overlap problem*

Generalization on the basis of overlap is a natural property of associative memories. Similar learned patterns each evoke elements of the other and tend to produce intermediate patterns. Such generalization is helpful if the overlap is associated with a cluster of patterns in the environment that merit similar responses, for example, oak trees. If it is the unique features of a pattern that are important at recall (for example, the scene of an accident), then generalization degrades performance. This is the overlap problem: a tendency for features to intrude into recall simply because they occur in patterns that overlap substantially with the required pattern.

The overlap problem can sometimes be alleviated by pre-processing of incoming information. Pattern classification, feature detection in sensory systems and 'codon' extraction in general neural processing (Marr 1969), can each operate to reduce overlap. However, at any level of processing there may remain overlapping patterns for which the distinctions are subtle but important. Consolidation from short- to long-term memory affords an opportunity to reduce the errors arising from this problem.

2. OPERATING PRINCIPLES

(a) *Auto-associative memory*

Auto-associative nets generally operate through excitatory associative connections that develop between co-active elements. They offer a simple model for pattern storage and retrieval. The standard paradigm for assessment adopted here is *seeded recall*, in which a partial pattern evokes a more complete version of the learned pattern (Willshaw *et al.* 1969), analogous to the phenomenon of gestalt 'completion' (Hebb 1949).

Mathematical analysis has been done on the behaviour of such nets, both fully connected (Longuet-Higgins *et al.* 1970; Lansner & Ekeberg 1985) and partially connected (Gardner-Medwin 1976; McLelland 1987). Partly connected nets are examined here. This avoids certain singular mathematical features of fully connected nets. Partial connectivity is also closer to real anatomy in most tissues and is more efficient (in terms of information stored per synapse) in large nets (Gardner-Medwin 1976).

(b) *Synaptic properties*

Doubly modifiable synapses are assumed to contribute an excitatory effect that is the product of temporary (T) and persistent (P) weights. A specific physiological mechanism might involve temporary changes in pre-synaptic transmitter release and persistent changes in dendritic spines (or vice versa): dendritic current is determined by both mechanisms in series.

The modification condition postulated by Hebb (1949) is assumed, whereby strengthening of excitatory synapses occurs only when there is both pre- and post-synaptic activity. This is simple, highly effective as an algorithm for one-trial learning in auto-associative nets, and related to known physiological mechanisms (see reviews by Bliss & Lynch (1988); Cotman & Monaghan (1988) and Gustaffson & Wigstrom (1988)). Figure 1 shows T and P changes in a small net.

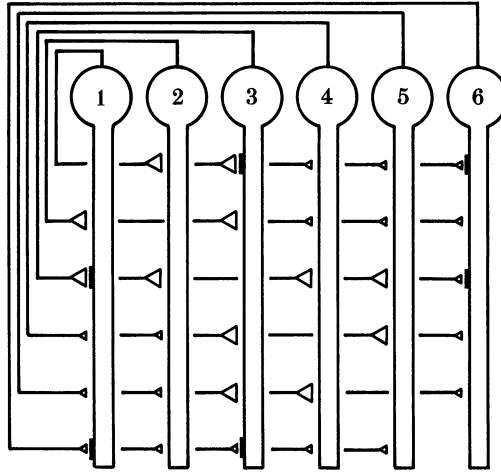


FIGURE 1. A small auto-associative net with doubly modifiable synapses. Temporary (T) synaptic strengthening is shown according to the Hebb rule for two patterns: 1, 2, 3 and 3, 4, 5 (large triangles). Persistent (P) strengthening is shown for pattern 1, 3, 6 (subsynaptic bars), but the corresponding T changes have since disappeared. Excitatory synaptic current is determined by the product of T and P , with the two physiological mechanisms operating in series.

(c) *Short-term memory*

Temporary (T) weights change from 0 to 1 whenever the Hebb rule is satisfied and revert to 0 after a short time. The multiplicative relation between T and P weights means that only synapses for which the Hebb rule has recently been satisfied will be effective at all. Recall performance normally depends on these synapses, so it is easy to see why recall is good for recent patterns: most of the synapses involved in earlier experience are ineffective and cause no interference (§4*b*). The effective synapses, with $T = 1$, generally have differing strengths due to their differing P weights. This degrades the performance somewhat by comparison with a conventional net that has uniform weights and the same short term experience; but such a net would have no long term memory. If just one or a small number of patterns have led to a few T changes, then it is always possible to obtain 100% recall performance for these patterns.

The timecourses of T and P changes are not made explicit in the model. They are expressed entirely in terms of the number of experienced patterns. The principles of the model may apply in different tissues with different timecourses, tailored to specific information processing requirements (§1*b*). The constraints are that T changes be long enough for consolidation of P changes to occur, but not so long that the rate of pattern presentation exceeds the short term memory capacity (§4*d*). If selective consolidation occurs during sleep (§5*c*), then T changes in appropriate tissues must last for hours or days to include such periods.

(d) *Consolidation*

The P weights have a small value to start with (relative to their maximum) and increase with many repetitions of the local Hebb conditions, or with repetitions

that are combined with a facilitating influence. The fact that only a small number of recent patterns are stored in T changes, and that these patterns can be accurately recalled, means that it is possible to re-evolve these patterns to cause increases of P . This is the process of consolidation.

Re-evocation of patterns may result from experience of a part of the pattern (seeded recall) or from sporadic random activity injected into the net. Random activity evokes one or other of the learned patterns because they alone are self-sustaining through their dense T connections. Similar consolidation of T to P synaptic weights was suggested in a previous paper (Gardner-Medwin 1969), but without the crucial multiplicative relation that allows short-term memory to generate high quality recall, to ensure changes at exactly the correct set of sites.

(e) *Long-term memory*

At first sight it might appear that the strengthened P weights will be of no value once the T changes have disappeared. After all, the high quality of short-term recall is achieved just because the synapses not involved in recent patterns are ineffective. Note however, that the required information is actually stored in the net. The P changes are the same as would be found in a conventional net with singly modifiable weights that had learned all the consolidated patterns. The problem is to gain access to this information.

There are at least three ways of benefitting from consolidated information. The net can *recognize* a pattern, that is, identify whether it has or has not been previously consolidated. This is relatively simple, because re-presentation enhances all the relevant T weights and enables the level of excitation within the pattern (i.e. the threshold at which the pattern is largely self-sustaining) to be used as a measure of consolidation.

Secondly, if the memory is *refreshed* by re-presentation of a consolidated pattern, recall will benefit during the period of T enhancement from the elevated P weights. This results in better short-term recall for consolidated patterns than fresh patterns (§4b).

Thirdly, it is still possible to *recall* a pattern from long-term memory, from presentation of a few elements, even if the pattern of T enhancements has disappeared. Most of the relevant synapses are ineffective ($T = 0$), and those that are effective through recent usage have probably only a random relation to the pattern to be recalled. In these circumstances straightforward recall is weak. The problem can be solved if T weights are raised temporarily at the synapses required for recall. This process is described as 'booting'.

(f) *Strategies for booting*

There are at least three strategies for booting.

1. The simplest is to set $T = 1$ at all synapses in the net. This turns the system (for the duration of T changes) into the equivalent of a conventional net, with all the long-term memories available for recall. The drawback is that all purely short-term memories are wiped out.

2. Less drastic but equally effective is to set $T = 1$ at just the synapses that project from cells that are active during the recall process. This affects all synapses

relevant to recall, but only a small fraction within the net as a whole. A physiological implementation might overcome the normal Hebb requirement for post- as well as pre-synaptic activity for T modifications, perhaps through strong activation of the seed cells or activation in combination with some sensitizing influence. The non-associative potentiation need only be short lasting, but it must interact with the T mechanism: synaptic strength must increase by a larger factor if the T weight is low than if it is high. In this respect it differs from the brief non-associative processes investigated by McNaughton (1983), which interact in a multiplicative fashion with longer-term processes.

3. Booting could be achieved by maintaining or repeating activity in the seed cells, while other patterns are generated within the net through normal usage on unrelated tasks or through deliberate recall of supposedly related patterns. The Hebb conditions would then ensure that T weights were enhanced for synapses from the seed cells to all cells active at any time during this usage. This will not generally increase all the relevant T weights, but it rapidly affects a majority. Suppose random patterns are employed, each involving activity in a fraction α (< 1) of the cells. After α^{-1} such patterns, 63% ($1 - \frac{1}{e}$) of the cells in the net will have been activated at least once, and correspondingly 63% of the T weights from the constantly active seed cells will have been enhanced. The required number of patterns is less than would saturate the short-term memory, as only a fraction α of the synapses within the net generally would be enhanced (α^2 for each of the α^{-1} patterns).

Strategy (2) is implemented in the simulations, since it is simple and complete. Strategy (3) is perhaps equally plausible. One consequence of (3) is that the seed pattern would become incorporated in other memories retained from the period of attempted recall. This is neither definitely deleterious nor implausible.

(g) *Optimization of consolidation*

With doubly modifiable synapses, the permanent engram need not be laid down at the time of learning. The initial period with high recall quality due to T changes gives a chance to optimize consolidation so as to improve the residual recall after disappearance of the T changes, when full interference from patterns in long-term memory will be experienced. Interference is caused by overlap between patterns; this leads to confusion and the intrusion of incorrect elements during recall. It will be shown how benefits can be gained through selective consolidation.

A pattern subject to consolidation may overlap with another in short-term memory, or with one previously consolidated and not in short-term memory. The former situation is the simpler, and is treated here. It might arise after initial learning at about the same time, or recent refreshment of either pattern. Overlap that involves old memories could be treated in much the same way, first refreshing the old memory with 'booted' recall (§3f). This might introduce complications through errors of recall, however, which are not considered.

There are at least two strategies to alleviate the overlap problem in an associative memory:

(1) Strengthen associations that are specific to each pattern, i.e. do not involve cells that are common to both. This improves the ratio of activation of the specifically correct cells to both incorrect and common cells throughout recall.

(2) Reduce the strength of associations between the common cells. This reduces the tendency for recall to be dominated by common cells, and is similar to the 'unlearning' process suggested for the elimination of parasitic recall states (Crick & Mitchison 1983; Hopfield *et al.* 1983).

The first strategy is more effective and is used in the simulations, though it is harder to implement. The problem is that there is no simple way to identify cells that are part of one pattern and not another. A two-stage process seems to be required (§3*g*) in which the first stage identifies common cells and then renders them non-functional ('fatigued') for the second stage. Both strategies require conditions that in some respects resemble sleep (§5*c* and Crick & Mitchison 1983).

3. MODEL DESCRIPTION

Principal definitions are summarized in table 1. The simulations were programmed in Microsoft 'C' on an IBM-PC compatible computer. Illustrated data points are all means for recall of at least five patterns.

TABLE 1. NOTATION AND VALUES USED IN SIMULATIONS

symbol	definition	value
N	number of cells	700
R	number of connections from each cell	500
W	number of cells active in each pattern	70
M	number of patterns learned	—
M_c	number learned & consolidated	—
M_{ST}	number learned since T weights set to 0	—
α	active fraction within a pattern	$W/N = 0.1$
a_i	activity of cell i	0 or 1
e_{ij}	excitation onto cell i from cell j	$T_{ij}P_{ij}a_j$
E_i	total excitation onto cell i	$\sum_j e_{ij}$
T_{ij}	temporary weight onto cell i from cell j	0 or 1
P_{ij}	persistent weight onto cell i from cell j	$p_{ij} + \beta$
p_{ij}	modifiable component of P weight	0 or 1 ^a
β	fixed component of P_{ij}	0.25
I_0	information within one pattern (equation 5)	328 bits
Q	quality of recall (equation 7)	0–100%

^a p_{ij} takes values 0, 1, 2, 3, etc., in simulations for §4*e*.

(a) Activity and excitation

The nets are partly connected, with 500 connections distributed randomly from each of 700 cells. Activity (a_i) is binary. Activation is either extrinsic or intrinsic. Extrinsic activation, in training and recall, occurs regardless of other conditions in the net. Intrinsic activation occurs when summed excitation (E_i) from other cells exceeds a set threshold.

Excitation E_i is given by:

$$E_i = \sum_j (T_{ij}P_{ij}a_j), \quad (1)$$

where summation is over all cells j that send projections to the cell i . T_{ij} and P_{ij} are the T and P weights onto cell i from cell j . Both have modifiable components that are binary (except P weights in §§3*g*, 4*e*). Increments from 0 to 1 occur when

the Hebb local associative condition ($a_i a_j = 1$) is satisfied along with other, enabling, conditions.

T changes are enabled during all learning trials. They do not occur during recall, except 'booted' recall (§3*f*), and they must be disabled during special procedures for selective consolidation (§3*g*).

P changes occur only if recall is combined with a facilitatory influence. P_{ij} has a fixed component $\beta = 0.25$, in addition to the modifiable component p_{ij} :

$$P_{ij} = p_{ij} + \beta. \quad (2)$$

The fixed bias is required to ensure that short-term memory works even in a naive network that has consolidated no patterns, or only a few. Without this, the multiplicative relation between T and P weights (equation 1) would preclude any excitation under these conditions. The choice of bias is justified in §4*c*.

(b) *Threshold control*

Threshold is variable, but uniform throughout the net. Occasionally it is set explicitly (§3*g*). During recall it is always controlled to achieve specific target numbers of active neurons. Physiological mechanisms for the necessary feedback are discussed elsewhere (Gardner-Medwin 1976).

(c) *Training and refreshment*

Patterns presented for learning were independent random sets of 70 active cells. In studies concerned with pattern overlap, such patterns were each paired with a second pattern having 52 randomly selected cells in common and 18 not in common (i.e. 74% overlap).

Training, equivalent to learning in short-term memory, sets $T_{ij} = 1$ at sites satisfying the Hebb condition. These T increments are maintained without decrement until a discrete time after training, when they all revert to zero. A variable number (M_{ST}) of patterns may be learned in short-term memory during this interval.

Refreshment is the relearning (with T changes) of patterns previously learned and consolidated.

(d) *Consolidation*

A learned pattern may or may not be consolidated, i.e. lead to changes of P weights. If not, then all trace of the pattern disappears when the T weights revert. Consolidation is taken to be all-or-none: if it occurs, p_{ij} is set to 1 wherever a T change has occurred, and remains high indefinitely. As will be seen (figure 4), 100% quality of short-term recall is readily achieved throughout the history of a net if there are no more than about five patterns stored in short-term memory at one time. Total, error free, consolidation is therefore consistent with implementations that might require multiple repetitions through recall of a learned pattern, or recall under special enabling conditions.

Selective consolidation to deal with the overlap problem (§3*g*) is achieved by allowing additional integer increases of p_{ij} (equation 2). For these simulations p_{ij} is multivalued and does not saturate. Increments occur each time the Hebb condition is satisfied for a new consolidated pattern.

(e) *Recall*

Recall is initiated by extrinsic activation of 10 cells selected at random from the required pattern, the 'seed'. For 'simple' recall, the threshold is set just low enough to recruit directly the correct total number of cells ($W = 70$): those having the greatest excitation from the seed. If more than 70 would be recruited, then a random selection is made from those with excitation exactly equal to threshold.

For 'progressive' recall (Gardner-Medwin 1976), cells are recruited iteratively, with thresholds just low enough at each stage to recruit at least five extra cells until a total of 70 are active. Cells (other than seed cells) are dropped from the active set if they cease to be among those with greatest excitation. Random selection is used, if necessary, at the last stage to give exactly 70.

With overlapping patterns, three seed cells are specific to the required pattern and seven common to the overlapping pattern, approximately the same proportions as in the full patterns.

(f) *Booted recall*

Recall from long-term memory generally requires enhancement of some T weights (§2f). 'Booting' is implemented by setting $T_{ij} = 1$ for all outputs from seed cells and from all cells recruited during progressive recall.

(g) *Selective consolidation*

This procedure temporarily fatigues the most readily activated cells in the net, and then strengthens connections between associated cells that remain. It is a two-stage process, operating with each pair of patterns in short-term memory after initial consolidation (§3d).

Stage one identifies cells receiving the greatest total weight of effective (T enhanced) inputs. These tend to be the cells common to the overlapping patterns. Starting with a random set of active cells, threshold is lowered to activate 50% more cells than in a normal pattern (105 cells). This recruits all cells from both patterns (plus some extra), and requires a threshold below 50. Threshold is then raised to 100, to retain only the cells with greatest excitation from this set. These are 'fatigued': the net is treated as if they did not exist for the subsequent stage two. Equivalent physiological implementations of fatigue could involve temporary inexcitability, or a weakening of afferent or efferent synapses for the cells maximally excited in stage one.

The large fluctuations of threshold during stage one are substantially above and below the level of excitation ($E_i = 62$) that is generated within an ordinary active consolidated pattern. E_i is calculated from equation (1), by setting $P_{ij} = 1.25$ (equation (2)) for 49 cells projecting, on average, to any one cell (i.e. five sevenths of the 69 other cells, taking account of partial connectivity). The fluctuations are greater than could be tolerated during normal recall and result, at low threshold, in hybrid patterns of cells that are only loosely associated.

In stage two, recall is again triggered by random activity, but under the tight threshold control of normal progressive recall. Random activity is first quenched to a few active cells by raising threshold; then progressive recall is used to recruit iteratively (without an extrinsic seed) to 25% of normal pattern size (18 cells).

These partial patterns tend to be the disjoint parts of one or other of the overlapping patterns, without the common cells. P_{ij} is incremented by one, wherever the Hebb rule is satisfied in these patterns and also $T_{ij} = 1$.

The T modifiability is disabled throughout this procedure. This is essential, at least for stage one, because of the spurious hybrid patterns. Four iterations of the two stages were performed, with recovery of 'fatigue' in between.

(h) *Assessment of recall quality*

The definition of recall quality, Q , is based on information theory principles. The information required to identify and correct errors in the recalled pattern is compared with that required to specify the pattern from scratch, and Q is defined as the percentage saving.

With W out of N active cells in a pattern, each cell has an *a priori* probability $\alpha = W/N$, the active fraction, of being part of the event (table 1). The information required to specify for each cell whether it is part of the event is the Shannon entropy $H(\alpha)$, where:

$$H(x) = -x \log_2(x) - (1-x) \log_2(1-x) \quad (\text{bits}). \quad (3)$$

The total information (I_0) required to specify the correct pattern from scratch is therefore:

$$I_0 = NH(\alpha). \quad (4)$$

Suppose that recall leads to a pattern with m cells missing and s cells spuriously activated, compared with the correct pattern. A total of $w (= W - m + s)$ cells will be active. The information (I_c) required to correct the errors consists of two components: information to identify the spurious cells amongst those that are active, and information to identify the missing cells amongst those that are inactive. Thus:

$$I_c = wH(s/w) + (N-w)H(m/(N-w)), \quad (5)$$

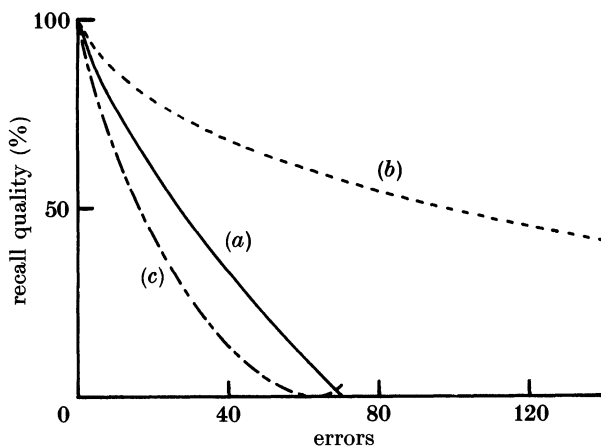


FIGURE 2. Recall quality plotted against number of errors of different types within a recalled pattern: (a) cells missing from the correct active set; (b) spuriously active cells in addition to the correct set; (c) incorrect cells active instead of cells in the correct set (i.e. correct total number of active cells).

where $H(x)$ is given by equation 3. Recall quality Q , the percentage savings, is then:

$$Q = 100\%(I_0 - I_c)/I_0. \quad (6)$$

Figure 2 shows quality as a function of the number of errors, for three conditions: a , missing cells; b , extra spurious cells and c , equal numbers missing and spurious. As the recall algorithms in this paper always generate the correct number of active cells (§3e), curve c applies. Initial seed patterns (with 10 correct cells active) have quality $Q_{SD} = 10.4\%$ (curve a).

4. RESULTS

(a) Basic memory performance

Figure 3 shows performance characteristics, without doubly modifiable synapses, for a conventional auto-associative memory. The theory behind this performance was analysed (without simulations) by Gardner-Medwin (1976). Data were obtained in the present model by enhancing only T weights for every pattern, with no reversion to zero. Recall quality declines as the number of trained patterns increases. Performance with simple recall deteriorates steadily from the start, but progressive recall maintains 100% performance for up to 50 learned patterns. This is similar to predicted behaviour (figure 4 & 8 of Gardner-Medwin 1976). Performance drops rapidly when the modified fraction of synapses (f in figure 3) exceeds about 50% ('saturation').

(b) Improvement of short-term memory

Recall in the conventional model (figure 3) is the same for new and old patterns. The ability to learn and recall fresh patterns declines with long-term experience. This decline is essentially completely overcome with doubly modifiable synapses.

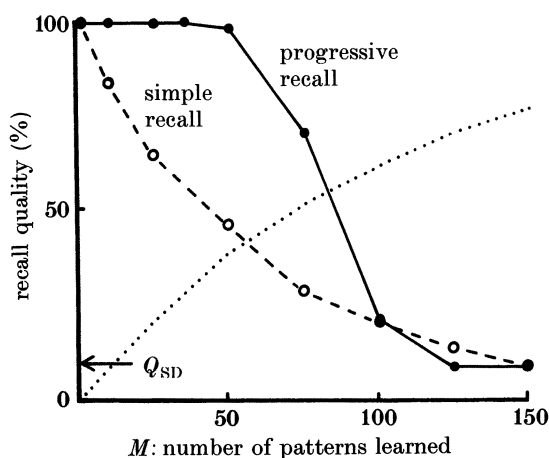


FIGURE 3. Recall quality (Q) in an auto-associative net with simple modifiable synapses having one binary strength parameter. Recall was elicited from a seed of 10 active cells, using simple and progressive recall procedures. Dotted line: fraction f of synapses in the net that have been modified. Q_{SD} : quality of seed pattern.

Short-term memory performance is shown in figure 4 for nets that have learned, consolidated and ‘forgotten’, in respect of T changes, up to 150 patterns ($M_c = 150$). At the start of the short term memory tests, these nets had the same pattern of P enhancements as the simple nets in figure 3, but everywhere $T_{ij} = 0$. Short-term recall was tested with five patterns in succession, enhancing T weights without decrement or consolidation. These patterns were either ‘new’ (i.e. never experienced before) or ‘refreshed’ (i.e. the same as five of the consolidated set). With all levels of prior training (M_c), the short-term recall with doubly modifiable synapses (figure 4) was better than for the conventional net (figure 3).

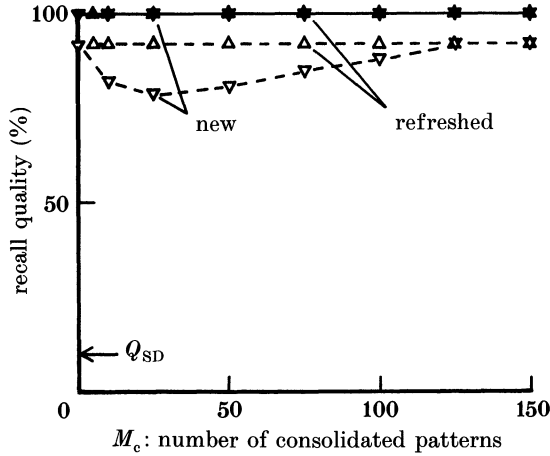


FIGURE 4. Short term memory in nets with doubly modifiable synapses. Recall quality is shown following T changes for five new patterns (∇, ∇) and five previously consolidated and refreshed patterns (\triangle, \triangle), as a function of the number (M_c) of previously consolidated patterns in long-term memory. Data are for simple (dashed line) and progressive (solid line) recall.

Progressive recall (filled symbols: figure 4) gave perfect recall throughout, i.e. at all levels of prior experience. Simple recall of new patterns (inverted triangles) gave performance that fell as the experience increased up to 25 patterns, then rose again. Even at its minimum, this performance was better than in the conventional net (figure 3). Simple recall with ‘refreshed’ patterns (upright triangles) was better, equivalent to a naive conventional net with experience of only five patterns.

The difference between new and refreshed patterns arises from the non-uniformity of the P weights connecting cells in the new patterns. The P weights are most uniform:

- (i) at the start when they are all small ($P_{ij} = \beta$);
- (ii) after much consolidation when most are large ($P_{ij} = 1 + \beta$). The non-uniformity can be expressed as the coefficient of variation of the P weights (standard deviation/mean). This passes through a maximum when the fraction f of the P weights that are enhanced is $\beta/(1+2\beta)$, or 0.17. This corresponds to 18 consolidated patterns (figure 3). Thus there is an initial decline in simple recall for new patterns, followed by an improvement (figure 4). The decline in discrimination between correct and incorrect cells is sufficiently small, however, that performance with progressive recall is uncompromised.

(c) Retrieval of long-term memories

Patterns for which T changes have reverted to zero require ‘booting’ for satisfactory recall (§§2*f*, 3*f*). Figure 5 compares recall without booting (open symbols) and with booting (filled symbols), in nets with different levels of consolidated experience (M_c). Simple recall (broken lines) and progressive recall (solid lines) are distinguished.

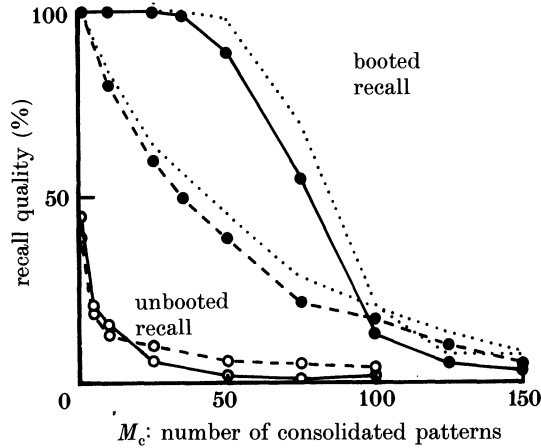


FIGURE 5. Long-term memory in nets with doubly modifiable synapses. Recall quality is shown for consolidated patterns that have been ‘forgotten’ in respect of T changes, without booting (open symbols) and with booting (filled symbols), plotted against the number (M_c) of such patterns. Data are for simple (dashed lined) and progressive (solid lines) recall. Dotted lines reproduce the recall quality obtained using singly modifiable synapses (figure 3). Data without booting was obtained with 20 unrelated patterns stored in short term memory. For definition of booting, see §3*f*.

Without booting there would be no recall at all, unless some synapses have $T_{ij} = 1$. The unbooted recall in figure 5 is for a situation with $T_{ij} = 1$ at about 18% of the synapses within the network, following T enhancements with 20 unrelated patterns. This is about the largest number of patterns that can be held in short-term memory consistent with reasonable ($> 75\%$) short-term recall (figure 6 and §4*d*). It is obviously not adequate to sustain satisfactory long-term retrieval without booting.

Retrieval is much better with booting (figure 5, filled symbols). Booted performance is almost as good as that shown in figure 3 for a net with simple modifiable synapses (reproduced as dotted lines in figure 5). It might be expected that performance would be *exactly* equivalent to a conventional net, since with booting $T_{ij} = 1$ for all active synapses, and performance is therefore determined simply by the pattern of P weights. The difference is due to the fixed component (β) in the P weights for doubly modifiable synapses (equation 2). During booting, synapses become temporarily effective that were never consolidated in the initial learning; these contribute some inappropriate excitation. The value $\beta = 0.25$ was chosen to be large enough that short-term memory is good throughout the history of a net (figure 4) and small enough that booted recall is nearly optimal (figure 5).

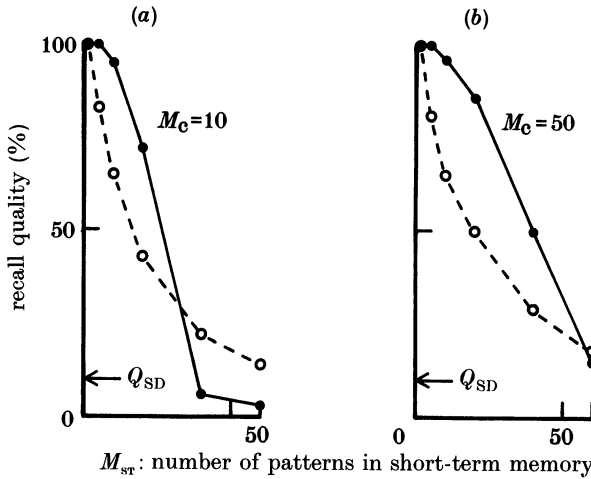


FIGURE 6. Short-term memory capacity. Recall quality Q is plotted against the number (M_{ST}) of new patterns in short term memory (T changes only), in nets with different amounts of long-term experience ($M_c = 10$ and $M_c = 50$: cf. figure 4). Data are for simple (dashed line) and progressive (solid line) recall. Memory capacity (number of patterns in short-term memory consistent with good recall) is less than for long-term memory (figure 5).

(d) *Comparison of short- and long-term capacity*

Although the quality of short-term recall for a small number of patterns is high (figure 4), the number of patterns that can be learned without serious interference is much less than for long-term memory. Figure 6 shows performance as a function of the number of new patterns stored in short-term memory (M_{ST}), for two levels of prior long-term experience ($M_c = 10$ and $M_c = 50$). As in figure 4, performance for up to five new patterns is excellent at both levels of experience. With more patterns, performance starts to drop. The 'capacity' of short term memory can be defined as the number of stored patterns consistent with, say, 90% quality of progressive recall. This is 12 with $M_c = 10$ and 15 with $M_c = 50$, compared with an equivalent long term capacity of 50 patterns with booted retrieval (figure 5), or 57 with a simple binary net (figure 3).

(e) *Recall of overlapping patterns*

The problem with overlap is illustrated in figure 7a. Long term recall performance is shown for patterns that occurred in pairs with 74% overlap. All the illustrated data are for progressive recall with booting. The circles show performance when ordinary consolidation procedures were used, with each pattern consolidated on its own to produce P increments. Performance fell rapidly, even with as few as 10 stored patterns (five pairs). This is much worse than for ordinary patterns (capacity: 50 patterns, §4d).

The recall quality for overlapping patterns fell rapidly to around 67%, which is the quality (Q_{COM} : figure 7a) for recall that identifies correctly the cells common to the two overlapping patterns but fails to discriminate between cells of the

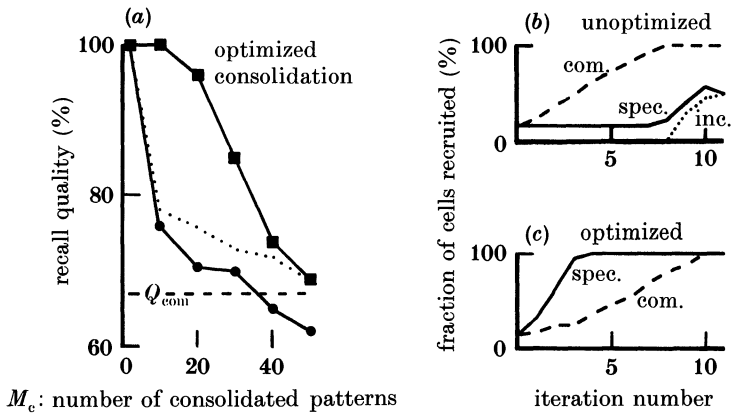


FIGURE 7. Long-term recall of overlapping patterns, with and without optimization by selective consolidation. Patterns were learned in pairs having 52 cells (74%) in common. (a): circles basic consolidation, equal P increments for every association in a learned pattern; dots, same, with saturation after one increment (binary p_{ij}); squares, as circles, with additional selective consolidation (§3g); M_c , number of consolidated, paired patterns; Q_{com} , quality with all common cells recruited, and equal numbers specific to correct and incorrect patterns. (b, c): recruitment sequence for common cells (dashed lines), specifically correct cells (solid lines) and incorrect cells (dotted lines) without (b) and with optimization (c); data as for $M_c = 20$ in (a). Means for $M_c > 10$ are for one from each of the first and last five pairs to have been learned.

specifically correct and incorrect patterns. The reason is shown in figure 7b, where the proportions of these three categories of cells that were recruited are shown for each stage of the iterative recall process. Apart from the 14% of common and specifically correct cells that were activated extrinsically from the start, all the first cells to be recruited are common cells. Only when these are all active are the more weakly associated specific cells recruited. By this time, however, incorrect cells are almost as strongly excited by the active common cells, and are recruited also.

Additional selective consolidation following the procedure for optimization (§3g) produced a marked improvement in recall of the overlapping patterns (squares in figure 7a). The procedure strengthens mainly the connections within non-overlapping parts of a pattern. Only one pattern benefitted at a time. Once the cells specific to one pattern were strongly associated, however, they tended to drop out through 'fatigue', along with common cells, in stage one of the procedure. This favoured the other pattern in subsequent iterations, so that four iterations led to benefit in both. Further iterations had little extra effect, because nearly all the cells dropped out in stage one of the procedure.

The number of stored overlapping patterns consistent with 90% quality of recall increased from six to 26 with the optimization procedure (figure 7a). This is less than the capacity of 50 for ordinary patterns (figure 5), but represents a marked improvement. The improvement was reflected in a different recruitment order for the classes of cells in progressive recall (figure 7c). Cells specific to the correct pattern were now mainly recruited *before* the common cells (contrast figure 7b).

Errors were sometimes introduced by the optimization procedure, whereby individual incorrect cells came to be strongly associated with the wrong pattern. As shown, however, the mean performance was much improved.

The data shown with full lines in figure 7*a* were obtained with many-valued P weights, making additional increments possible (§3*d*). One consequence is that connections between common cells were doubly strong ($p_{ij} = 2$) after initial consolidation. This effect contributes to early recruitment of common cells (figure 7*b*) and can be dealt with more simply than through an algorithm for selective consolidation, by reverting to binary weights. Only a slight improvement in performance results, however (dots in figure 7*a*). A statistical preference for recruitment of common cells remains even with binary weights, and there is little change in the ratio of excitation experienced by correct and incorrect cells.

5. DISCUSSION

(a) *Relation to synaptic properties*

An essential feature of the doubly modifiable synapses is the multiplicative relation between T and P synaptic weights. The T modifications must essentially switch the synapses on and off if full advantage is to be gained for short-term memory (§4*b*). A multiplicative relation has been demonstrated in the rat hippocampus by McNaughton (1983) between associative long-term potentiation, which lasts hours or days, and the briefer (non-associative) phenomenon of augmentation. For the ideas of the current paper to be correct, it must also exist between associative phenomena of different timecourses.

Timecourses of T and P changes may depend on the situation in different tissues (§2*c*). The P changes may require enabling influences during consolidation (§2*d*), whereas both changes may need to be disabled during sleep (§§3*g*, 5*c*).

(b) *Relation to cognitive systems*

A simple auto-associative net, though efficient for the storage of patterns in a content-addressable form, would be simplistic on its own as a model of a cognitive memory system. A realistic memory must involve other forms of plasticity and other algorithms, for example error-correction algorithms that are well suited for learning to classify and generate patterns. Auto-associative storage may simply be one feature of a memory system, operating at different levels.

Long-term memory tends to be retained at higher levels of processing (for review see Baddeley (1986)). This does not mean that the present model, with short and long-term memory in a single network, is wrong. The tendency for long-term memory to be more evident at higher processing levels may simply reflect different decay parameters for persistent traces or different interference at different levels.

Several important features of memory systems are duplicated in the results demonstrated here for a simple net: high quality short-term recall, low short-term capacity, sensitivity of short-term memory and consolidation (but not the recall of long-term memories) to disturbances that cause extremes of activity, ability to recall old memories in a slow painstaking fashion ('booting'), ready access to old memories that have been recently recalled in this fashion, and particularly good recall for old consolidated patterns that have been recently refreshed.

Consolidation is envisaged as due to repeated recall of patterns or parts of patterns (§§2*d*, 3*d*). The simulations employ total consolidation, with P weights elevated at all sites where T weights were set during initial learning. In real nervous systems, consolidation might occur with several patterns stored through T changes together, causing, if there were too many patterns, a loss of recall accuracy. It might also be graded and occur at different strengths for different patterns, or affect only a proportion of the relevant synapses. In these respects the analysis is simplified to demonstrate basic principles.

(c) *Relation to sleep*

One advantage of a memory with consolidation is the opportunity it affords to take account of other stored patterns in determining the persistent engram. The procedure for selective consolidation of the connections, improving recall of overlapping patterns (§4*e*), has several parallels to known facts about sleep.

The procedure requires isolation from sensory and motor interactions, since hybrid and partial patterns occur that were never experienced and must not be remembered or interpreted as real. It requires two alternating states, which resemble slow wave sleep (cf. stage one) and paradoxical sleep (cf. stage two). Slow-wave sleep is associated with large swings of neuronal threshold (Evarts 1964), as required in stage one, and a depression of both short-term memory (Stones 1977) and hippocampal long-term potentiation (Leonard *et al.* 1987).

Stage two must follow stage one, and involves recall under tight threshold control similar to waking. These are characteristics of paradoxical sleep (Aserinsky & Kleitman 1953; Evarts 1964). Intermittent activity is required to initiate recall. Such activity occurs in the cat visual pathway during paradoxical sleep (Bizzi & Brooks 1963), and projects to the cortex (Gardner-Medwin 1974).

These parallels can be no more than suggestive of a true relation between sleep and selective consolidation. In critical areas there are no data, for example relevant to the carry over of some sort of 'fatigue' of highly active cells from stage one to stage two (§3*g*), or the importance of sleep for the long-term handling of overlapping memories.

I thank M. Recce for discussions and programming advice and many colleagues for comments and suggestions.

REFERENCES

- Aserinsky E. & Kleitman N. 1953 Regularly occurring periods of eye motility and concomitant phenomena during sleep. *Science, Wash.* **118**, 273-374.
- Baddeley, A. 1986 *Working memory*. Oxford: Clarendon Press.
- Bizzi, E. & Brooks, D. C. 1963 Functional connections between pontine reticular formation and lateral geniculate nucleus during deep sleep. *Archs. ital. Biol.* **101**, 666-680.
- Bliss, T. V. P. & Lynch, M. A. 1988 Long-term potentiation of synaptic transmission in the hippocampus: properties and mechanisms. In *Long-term potentiation: from biophysics to behavior* (ed. P. W. Landfield & S. A. Deadwyler), pp. 3-72. New York: Alan R. Liss, Inc.
- Cotman, C. W. & Monaghan, D. T. 1988 Excitatory amino acid neurotransmission: NMDA receptors and Hebb-type synaptic plasticity. *A. Rev. Neurosci.* **11**, 61-80.
- Craik, F. I. M. & Lockhart, R. S. 1972 Levels of processing: a framework for memory research. *J. verb. Learn. verb. Behav.* **11**, 671-684.

- Crick, F. & Mitchison, G. 1983 The function of dream sleep. *Nature, Lond.* **304**, 111–114.
- Evarts, E. V. 1964 Temporal patterns of discharge of pyramidal tract neurons during sleep and waking in the monkey. *J. Neurophysiol.* **27**, 152–171.
- Gardner-Medwin, A. R. 1969 Modifiable synapses necessary for learning. *Nature, Lond.* **223**, 916–918.
- Gardner-Medwin, A. R. 1974 Optic radiation activity during sleep and waking. *Expl Neurol.* **43**, 314–329.
- Gardner-Medwin, A. R. 1976 The recall of events through the learning of associations between their parts. *Proc. R. Soc. Lond. B* **194**, 375–402.
- Gustafsson, B. & Wigstrom, H. 1988 Physiological mechanisms underlying long-term potentiation. *Trends Neurosci.* **11**, 156–162.
- Hebb, D. O. 1949 *The organization of behavior*. New York: John Wiley.
- Hinton, G. E. & Plaut, D. C. 1987 Using fast weights to deblur old memories. In *Proc. 9th A. Cog. Sci. Soc. Conf.* Erlbaum.
- Hopfield, J. J., Feinstein, D. I. & Palmer, R. G. 1983 'Unlearning' has a stabilising effect in collective memories. *Nature, Lond.* **304**, 158–159.
- Kohonen, T. 1984 Self-organization and associative memory. Berlin: Springer-Verlag.
- Lansner, A. & Ekeberg, O. 1985 Reliability and speed of recall in an associative network. *IEEE Trans. PAMI* **7**, 490–498.
- Leonard, B. J., McNaughton, B. L. & Barnes, C. A. 1987 Suppression of hippocampal synaptic plasticity during slow-wave sleep. *Brain Res.* **425**, 174–177.
- Longuet-Higgins, H. C., Willshaw, D. J. & Bunemann, O. P. 1970 Theories of associative recall. *Q. Rev. Biophys.* **3**, 223–244.
- Marr, D. 1969 A theory of cerebellar cortex. *J. Physiol.* **202**, 437–470.
- McClelland, J. L. 1987 Resource requirements of standard and programmable nets. In *Parallel distributed processing* (ed. Rumelhart, D. E. *et al.*), vol. 1, pp. 460–487. Cambridge, Massachusetts: MIT Press.
- McNaughton, B. L. 1983 Activity dependent modulation of hippocampal synaptic efficiency: some implications for memory processes. In *Neurobiology of the hippocampus* (ed. W. Seifert), pp. 233–252. London: Academic Press.
- Palm, G. 1988 On the asymptotic information storage capacity of neural networks. In *Neural computers* (ed. R. Eckmiller & C. V. D. Malsburg), pp. 271–280. Berlin: Springer-Verlag.
- Stones, M. J. 1977 Memory performance after arousal from different sleep stages. *Br. J. Psychol.* **68**, 177–181.
- Willshaw, D. J., Buneman, O. P. & Higgins, H. C. 1969 Non-holographic associative memory. *Nature, Lond.* **222**, 960–962.