ELSEVIER

Behavioural Brain Research 69 (1995) 167–175

BEHAVIOURAL BRAIN RESEARCH

Research report

Possible mechanisms for reducing memory confusion during sleep

A.R. Gardner-Medwin*, S. Kaul

Department of Physiology, University College London, London, WC1E 6BT, UK

Received 1 September 1995; accepted 1 December 1995

Abstract

Confusion readily occurs in memory processes between patterns that overlap substantially. Possible mechanisms are considered that might operate in an automatic manner to reduce such confusion. One such mechanism is the recall of patterns in a distorted way, so that they are enriched with greater activity in distinctive elements of experienced patterns than in overlapping elements. Selective consolidation based on such enriched patterns would reduce confusion in long-term recall and might benefit discrimination learning. It is shown how automatic algorithms could achieve this through a process with two phases. In the first phase, somewhat analogous to slow-wave sleep, it is necessary for the normal tendency of the nervous system to learn correlations of associated activity to be disabled. The second phase must occur while the cells most active in the first phase are relatively inexcitable. Enriched patterns would be generated during this phase through recall, which might be triggered by bursts of activity such as occur in rapid eye movement sleep. Selective consolidation would take place during the second phase. If such processes do occur in the nervous system, it seems likely that they would have evolved to occur during sleep.

Key words: Sleep; Memory; Recall; Consolidation; Overlap

1. Introduction

It is a biological fact that most animals sleep. We should therefore expect that many physiological functions will have adapted through evolution to take advantage of the conditions during sleep. A valid approach to sleep research may be to mimic this process: to ask how known limitations on function might benefit through processes that could more readily take place during sleep than during waking. The answers may not correspond to the mechanisms that have evolved. However, the special conditions that may be required for benefits to take place and the specific predictions may stimulate questions for experimental research. We adopt this approach here, in relation to a problem we describe broadly as that of *confusion* in memory mechanisms.

The paper presents an overview of the issues, some of which are set out in more technical detail in previous publications [11,12]. We consider first what we mean by confusion, how it affects biological fitness and the ways that there are in principle to get round the problems. We then consider how information processing algorithms could be implemented to diminish confusion and how such implementation could benefit from the existence of a sleep state.

A particularly interesting conclusion is that maximum benefit seems to require an interplay between two radically different and alternating brain states, with different characteristics. We refer to these as Phases 1 and 2 of the processes leading to benefit. The corresponding functional states of the brain would resemble, in at least some superficial respects, the states of slow-wave sleep (SWS) and rapid eye movement sleep. The required interaction between the alternating phases is not known to exist between the states of sleep, but would not be readily evident unless specifically sought experimentally. In essence, the suggestion is that memory consolidation takes place in association with recall in Phase 2 (analogous to REM sleep) and is rendered particularly beneficial by transient consequences of specific activity that occurred in the preceding Phase 1 (analogous to SWS). The after-effects of Phase 1 would most simply involve a short-lived diminution of excitability or efficacy in cells that were strongly active during this phase. The conditions in Phase 1 that lead to this selective dropout must not themselves be re-

* Corresponding author. Fax: +44 1 71 3837005; e-mail: a.gardner-medwin@ucl.ac.uk

SSDI 0166-4328(95)00011-9

membered or consolidated. Their beneficial effect is to enrich the subsequent recall and consolidation through greater relative activation of cells that do not contribute to confusion.

The hypotheses arising from this work are consistent with the more general hypotheses put forward by Giuditta et al. [14] and Clark et al. [3], that a period of unlearning of some sort might take place in SWS, followed by consolidation in REM sleep. However, the effects of SWS, for the current proposals, need only be transient, since they serve merely to condition the system for the subsequent consolidation period in REM sleep. The transient effects of SWS could resemble the kind of long-lasting after-effect proposed by Crick and Mitchison after REM sleep [4,5]; however, the role of REM sleep in the present proposals is opposite to that suggested in their work. The synapses that are selectively strengthened in REM sleep with the present algorithms give a greater long-term advantage, it is argued, than the simpler single-stage unlearning algorithms.

2. Generalisation, overlap and confusion

The problem we address as *confusion* is a failure to distinguish patterns that are similar, but that have important differences. Confusion is essentially the same phenomenon as what is less disparagingly called *generalisation* when differences between patterns happen to be unimportant. Generalisation is a well-known and natural characteristic of neural network solutions to pattern classification problems [21]. Associations learned with one set of input patterns tend to transfer readily to testing situations in which the patterns are similar to those in training, but not identical. This can be a valuable characteristic of biological systems, as for example when an animal correctly generalises responses it has learned in a limited set of either dangerous or beneficial situations. Generalisation becomes undesirable and referred to as confusion, when small differences between input patterns are crucial: for example, when we ascribe the same name to two people who look alike.

The reason that generalisation is so conspicuous a characteristic of neural network behaviour is that the activity of each element or cell in a network usually depends on summed influences from a large number of other elements. The characteristics of the sum are relatively insensitive to differences in any small number of the individual influences. Two patterns of activity in which most of the active cells are identical (i.e., which *overlap* a lot) tend to generate similar summed influences on other cells. This is not an absolute rule of course, since the cells that are differently active may have profound effects, outweighing the influence of the common cells. This can happen particularly where a small number of cells exert powerful inhibition on output cells. However, generalisation is a natural feature of situations where there are many similar summed influences.

It is necessary to introduce some terminology and simplifying assumptions for our discussion. We are often concerned with pairs of overlapping patterns of activity such as P1 and P2 in Fig. 1. These patterns are treated as being within a largely homogeneous population of N cells that contain representations of sensory events. For simplicity, activity of individual cells is taken to be binary (i.e., they are active or not active, without graded levels of activity). This is strictly unrealistic in relation to the nervous system, but it simplifies discussion of the problems without, so far as we can see, in any way generating them. The fraction of cells (α) active in one pattern we refer to as the *activity ratio* for that pattern. The fraction of the cells (β) in one pattern that are also active in the second pattern we refer to as the *overlap fraction*. The cells that are active in both patterns are *common* cells with respect to the two patterns and those that are active in only one of the two patterns are *distinct* cells. There are two sets of distinct cells when considering P1 and P2, designated D1 and D2 in Fig. 1. Cells in D2 constitute *intrusion errors* from P2 if they become active when P1 is appropriate and vice versa. The rest of the population (the majority of the cells when α is

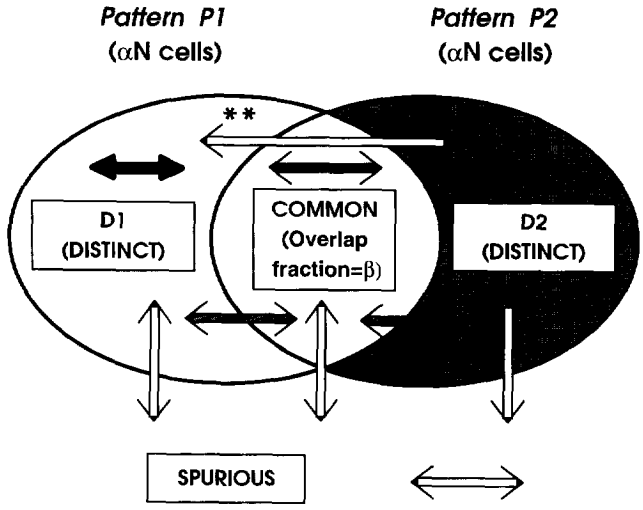

Fig. 1. The nomenclature and cell groups used in relation to overlapping sets of binary active cells, forming patterns of activity P1 and P2. The total number of cells is N, the activity ratio for any one pattern is α and the fractional overlap between patterns is β. The different categories of connections in relation to the problem of confusion between P1 and P2 are indicated by arrows. Excitatory connections that need to be selectively strengthened to diminish confusion in recall of P1 or P2 are shown black; those that assist recall but also contribute to confusion are shown as grey; other connections that will only be strengthened through other experience, neither P1 nor P2, are shown white. It is particularly important during procedures for selective consolidation not to strengthen connections between D1 and D2, marked **.

small) have no involvement in P1 or P2 and are referred to as *spurious* cells in relation to the confusion problems associated with P1 and P2.

Generalisation and confusion may be altered by recoding information to a new representation with altered overlap characteristics. Consider an example in the visual system. Identical geometrical patterns of light at different sites on the retina may activate wholly distinct sets of cells in the eye. At the stage of representation within the visual cortex it is found that overlap has been introduced. The cortical representations of visual stimuli include cells that are sensitive to features of shape at the retina, substantially independent of position. Shape is an important type of similarity between stimuli, sometimes more important than position. It is easy to see how overlap between representations of stimuli with similar shape may introduce valuable generalisation, while overlap based on position would contribute to confusion between stimuli of different shapes.

The difference between desirable generalisation and undesirable confusion rests on which differences between stimuli are *important*. This means, of course, important for the *animal*. The biological function of the nervous system is to generate actions appropriate to circumstances, best tailored to ensure survival for the animal and its species. *Importance* means relevance to such choices of action. An ideal representation of the environment, based on such considerations, would be one in which overlap occurs when and only when different stimuli have similar implications for action. Such an ideal state of affairs cannot exist for all stimuli, however. Animals are bound to experience stimuli from time to time that differ substantially from others in their implications, but little in their representations. The problem of confusion arises acutely in such circumstances and poses a challenge in how it may be minimised.

3. How might sleep aid in reducing confusion?

We distinguish two fundamentally different procedures to reduce confusion. The first is to alter representations of stimuli to reduce overlap fractions. The second is to retain overlapping representations but to alter selectively the strengths of synaptic connections involving the distinct and common parts of overlapping patterns. The first is the more obvious approach and indeed is one of the consequences of our natural waking reactions to confusion situations.

Consider an everyday example. Suppose you repeatedly leave the house having mistakenly picked up your spouse's keys instead of your own. A natural sequence of responses might be the following:

(i) Worry about the problem.
(ii) Pay more attention to the keys, thus increasing the number of attributes of them that are represented in the brain at the time a decision is made.
(iii) Identify distinct aspects of the two sets of keys and enhance the attention paid to these aspects.
(iv) If all else fails, label the two sets more distinctly.

One effect of these actions is to alter the representation of the keys generated in the brain when the keys are seen. The fourth action (probably the first action of a *sensible* person!) does this in a particularly direct manner. However, in considering the general neurobiological problem of confusion we should regard this as cheating, since it is often not possible to influence the external nature of confused stimuli. Steps (ii) and (iii) alter both the number of features represented and the balance between distinct and common features. The benefits will be similar to those arising in multi-layer artificial networks [22] where the selection and adjustment of intermediate feature-detecting units generates better representations for a particular task.

Step (i) in the table of responses (worry) is not included to be flippant but as a reminder that this simple instance of confusion is just the kind of thing that can have a surprisingly profound influence on behaviour. In particular, it is just the sort of problem that might influence sleep experience. There may be processes taking place during both sleep and waking that we are not aware of. It has been suggested by Marr [20] that alterations of stimulus representation may sometimes with advantage occur during sleep. Selective adjustment of synaptic weights (without altering overlapping representations) has also been proposed [4] as a mechanism to reduce the tendency of network dynamics to become dominated by readily elicited states.

The thesis advanced here is that it is in principle possible during a sleep-like state to generate recalled patterns to include largely the distinct elements of normal overlapping representations (either D1 or D2, but not both together, in Fig. 1). Memory consolidation based on these abnormal patterns, enriched with distinct features, may diminish subsequent confusion when tasks involve the full patterns.

4. Recall through auto-associative excitation

Recall or internal re-creation of a representation of experienced events, is a conspicuous feature of our own use of memory. We know this from introspection in ourselves and through language or artistic communication. We experience forms of recall during both waking and sleep. We have few means of studying recall directly in non-human

animals, though it seems likely that it occurs in some manner in many species. The ability to identify patterns as familiar (*recognition memory*) requires similar stored information and is better established in experimental animals [7].

The recall of a nearly accurate representation of an experience from a trigger involving a few active features and the settling of similar patterns (or even random patterns) into familiar prototypes are well-known examples of behaviour that can readily be produced in neural networks through the process of auto-association [10,11,17,13,24]. Mutually excitatory interactions between cells, if strengthened during occurrence of patterns through the type of mechanism proposed by Hebb [15], tend to support preferentially the recurrence of these patterns. This requires only one stimulus presentation, as in one-trial episodic learning. There is no scope for later error-correction, as can occur with repetitive training, since the *correct* pattern is not available for repetition. One cannot in general *replay* a face that has been seen once, to establish if recall of it is correct.

Auto-associative recall requires sparse coding of representations and a certain minimum density of connections between cells to be efficient. The constraints for handling randomly related patterns [10,11] are given approximately by the following relationships:
$a < \approx 1/\sqrt{M}$ and $R > \approx 30/a$, where a is the activity ratio (Fig. 1), M the number of patterns to be learned and R the number of other cells within the population (irrespective of size) to which each cell is connected. A larger a than the limit leads to poor performance, while a well below the limit means that an extravagant number of connections is required, as indicated by the second constraint.

When there is overlap between experienced patterns, recall tends to generate hybrid intermediates between the overlapping patterns. Recall of pattern P1 in Fig. 1 would readily lead to activation of cells within D2. We call these *intrusion errors* from P2. Intrusion errors from D1 would similarly occur during recall of P2. In the extreme of total confusion, recall of either P1 or P2 might lead to the same hybrid pattern, including elements of each.

Consider an example of overlapping episodic memories. Suppose you visit two similar sites on an excursion, say Buckingham Palace and Hampton Court. When recalling one of these, you may introduce intrusion errors from the other, with the likelihood increasing with the degree of similarity of your experiences in the two sites. If recall takes place after several years, residual long-term memory will be less reliable but may still include accurate recall of many details. Intrusion errors may be more common and there may even be total confusion in the sense that you cannot discriminate the two memories and may not even be aware that there were two different sites.

Confusion (or generalisation) errors always amount to a degradation of episodic recall and they must reduce whatever value it may have. Short of revisiting the sites, how can confusion be reduced? A diligent tourist might make a conscious effort, in the days after learning, to reduce the likelihood of confusions in long-term memory. A possible strategy would be to recall the separate sites (using the detailed and relatively reliable transient memory still available), to identify distinct features of each site and to think of associative links amongst such features. Not many of us take time to do this kind of thing, at least not while we are awake. Could it be done automatically, as part of the process of memory consolidation?

5. Reduction of confusion in recall through selective consolidation

Confusion in long-term recall can be reduced if the relative strengths of associations underlying long-term memory for overlapping patterns are adjusted selectively [11,12]. Fig. 1 shows the four-cell categories that enter into a confusion problem involving two specific patterns. These are the *distinct* cells (D1 and D2), the *common* cells and *spurious* cells that should not be active in either pattern. There are many classes of connection amongst these categories, shown as arrows in Fig. 1. Each class plays a different role in recall and confusion.

The black arrows are connections that aid correct recall of P1 and P2 and do not contribute to intrusion errors between P1 and P2. Selective enhancement of these is beneficial. The grey arrows are connections that assist recall of P1 and P2, but also contribute to intrusion errors. The COMMON-to-D1 and COMMON-to-D2 connections do this directly. For example, after recall of P1 the COMMON-to-D2 connections may provide activation of D2 cells above threshold. The COMMON-to-COMMON and DISTINCT-to-COMMON connections contribute to intrusion errors less directly: the problem arises early in a progressive recall process [10] triggered by a set of active cells that includes cells from D1. Strong COMMON-to-COMMON and DISTINCT-to-COMMON connections lead to early recruitment of common cells. These are not errors, since they belong to P1, but they tend to lead to intrusion errors. A preponderance of active common cells early in recall, when the rest of the D1 cells have yet to be recruited, results in almost as much activation onto D2 cells as onto the desired D1 cells. This can lead to a hybrid pattern or to a high probability of settling eventually into a stable recall of P2 rather than P1.

To achieve the best long-term recall it is necessary to have strong connections corresponding to both the black and the grey arrows in Fig. 1. The black connections,

however, need to be stronger. The other connections (white) are not useful in recall of P1 or P2, though they may be important in recall of other patterns; they are best left unchanged in procedures to aid recall of P1 and P2. It is particularly important that those marked ** should not be strengthened, since they would directly contribute to intrusion errors.

The desired discrimination between the black and grey connections in Fig. 1 poses an interesting theoretical challenge: how can the nervous system identify these connections and treat them selectively? Even with good transient memory mechanisms for P1 and P2, this is not straightforward. Individual synapses can be identified by their pre- and post-synaptic cells. With Hebbian modification conditions they are strengthened when both are active together. Thus the joint set of black and grey connections will be selected and strengthened with such mechanisms whenever there is separate recall of P1 and P2. Selective strengthening of the black connections would require that the partial patterns (just D1 or just D2) be activated on their own. Alternatively and providing only a partial solution, the COMMON-to-COMMON connections (though not the COMMON-to-DISTINCT and DISTINCT-to-COMMON connections) could in principle be selected and weakened relative to the others by activation of just the common cells.

It is simpler to see how the nervous system might selectively activate the common cells than the distinct cells. If recall of P1 is allowed to proceed under conditions with little inhibition and generally low neural thresholds, it will readily lead to recruitment of *all* the cells in either P1 or P2, forming a hybrid pattern (Fig. 2). If this state is rapidly followed by a sharp increase of thresholds, the cells that remain active longest will be just the common cells, since these have the greatest number of afferent connections strengthened by experience of P1 and P2. Thus with alternations of low and high threshold beyond the limits that are normally desirable for accurate recall of experienced patterns, the common cells will be those that are activated most and that remain active with the greatest inhibition (Fig. 2, Phase 1). Weakening of the connections between these cells with an associative 'unlearning' mechanism is essentially the kind of mechanism proposed by Crick and Mitchison [4] and would achieve part but not all of what is required for differential consolidation in the present context. It must be noted that if the nervous system is to operate in this way it is essential that the hybrid patterns (including, e.g., all cells of either P1 or P2) should not lead to associative strengthening of active connections, since this would strengthen the connections marked ** in Fig. 1 that would highly effectively cause intrusion errors. If the procedure takes place, it must take place with normal learning mechanisms disabled.

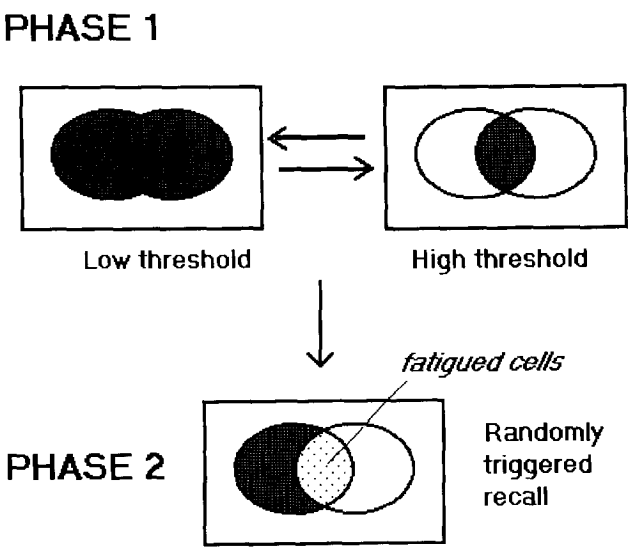

Fig. 2. A 2-phase procedure for activating the distinct cells (crescent shaped zones in the diagram) that are not part of the overlap between a pair of patterns. Phase 1 consists of alternate high and low threshold levels imposed on the cells, leading to hybrid patterns (at low threshold) and to patterns with the greatest number of strong afferent connections from these hybrid patterns (generally the common cells) at high threshold. The cells still active at high threshold are rendered inactive or inexcitable following Phase 1–2. Phase 2 is a period in which randomly triggered recall of patterns occurs, leading to activation of sets of cells that are strongly bound by associative connections, but that were not amongst the most readily activated in Phase 1. Associative strengthening of connections must not take place during activation of the hybrid patterns in Phase 1.

The distinct cells in P1 and P2 are not so easily activated on their own as are the common cells. There does not seem to be any direct manoeuvre that would render them the most easily activated cells. However, prior activation of just the common cells could make this possible if, as a result, the common cells somehow become difficult to activate. This is the basis of the 2-phase mechanism proposed for optimal selective consolidation [11] (Fig. 2). Phase 1 requires alternating high and low threshold conditions, without strengthening or consolidation of synapses between active cells. This leads to selective fatigue of the cells that are most active in the high threshold state of phase 1, lasting into phase 2. Normal recall processes become effective in Phase 2, triggered for example by random activation [8] under conditions of tight threshold control. This ensures that self-sustaining sets of cells become active, bound by strong excitatory connections: patterns such as D1 or D2 are activated, but not both together. Consolidation of the connections that are active both pre- and post-synaptically could then occur in this phase. Depending on the degree of fatigue of the common cells, the patterns forming the basis for this selective consolidation will be either exclusively the distinct cells or enriched with greater activation of these cells. The nature

of the fatigue process is not very critical so long as it reduces the recruitment of common cells: it could be a temporary elevation of threshold or a reduction of synaptic efficacy of synapses onto these cells.

Selective consolidation to reduce confusion in long-term recall requires a transient memory that lasts through the consolidation process. Such memory mechanisms exist on many different time scales [1], but their mechanisms are not well understood. A specific model of transient and long-term memory, where both changes reside in the same set of synapses, has been used [11] to assess the effect of the selective consolidation procedures on long-term confusion. Fig. 3 shows recall quality for pairs of patterns that overlapped by 74%, with and without selective consolidation. The unselective condition (circles) led to uniform increments of synaptic weight for each connection in the experienced patterns: recall quality fell rapidly as more patterns were stored, to levels corresponding to total confusion between the paired patterns. With selective consolidation (square symbols) recall was markedly improved. The selective procedure (Fig. 2) was repeated after learning of each pair of patterns until it was no longer selective: eventually the activation of all cells in Phase 1 becomes approximately uniform and Phase 2 ceases to be selective.

Some of the improvements in recall evident in Fig. 3 might have been due merely to equalising the connection strengths within the distinct and common zones of overlapping patterns. With unselective consolidation (circles), double usage of connections in the common zone produced double strengthening. Eliminating this effect in a more simple manner by using binary synapses in the simulations (dashed line) produced, however, much less improvement than did the full algorithm for selective consolidation.

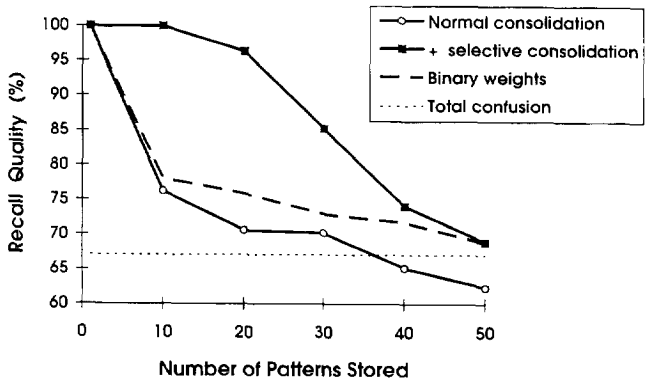

Fig. 3. The effect of selective consolidation on recall performance with overlapping patterns. Up to 50 patterns were experienced and stored, each consisting of 70 active cells out of 700. Each pattern had a twin amongst the set, with which it had 74% overlap (52 cells). Recall quality is plotted as a function of the number of patterns stored. Conditions were: normal consolidation consisting of equal increments of synaptic weight for each association experienced (circles); consolidation using binary weights for which all experienced associations have equal weight (broken line); normal consolidation plus selective consolidation that preferentially increased weights for connections between distinct cells in each pattern according to an automatic algorithm described in the text and in more detail in the original paper [11]. Recall quality of 67% corresponds to correct recall of common cells but total confusion (i.e., no discrimination) between distinct cells of the pattern to be recalled and its overlapping twin. Recall with normal consolidation (using either graded or binary weights) fell rapidly to around this level, while performance with selective consolidation was much better.

6. Might selective consolidation occur during sleep?

Selective recall of patterns, enriched with distinctive features, requires special conditions. Phase 1 leading to this involves the generation of hybrid patterns in which associations of activity must not be learned. If this is to be achieved (e.g., through neuromodulatory influences on synaptic plasticity) it requires that the relevant parts of the nervous system may not, at the same time, be involved in learning associations due to external stimulation. Sleep periods, when animals are cut off from external stimulation, would seem the best times to implement such a mechanism.

Phase 1 requires synchronous fluctuations of neuronal thresholds in whole populations of cells. Fluctuations that superficially resemble such conditions are known to occur in slow-wave sleep [6]. Memory and synaptic plasticity are also known to be impaired in slow-wave sleep [23,19,2,18]. Thus it is tempting, if speculative, to suggest a parallel between SWS and Phase 1. This phase must occur before and is in essence only the prelude to Phase 2 in which enriched patterns are generated and consolidated. Phase 2 requires recall to take place with tight threshold control, similar to that which presumably normally occurs in waking, to ensure that only sets of cells bound together through learned experience become activated. Consolidation of connections must occur in Phase 2, which might occur under waking or sleeping conditions. However, the activity patterns will be abnormal because of the after-effects of the prior Phase 1. If an animal is awake in Phase 2, it will be in an abnormal state in which elements of normal recall are hard to elicit. It is tempting to suggest that Phase 2 might occur specifically in REM sleep, conditioned by the prior SWS and with recall triggered by the bursts of activity associated with PGO waves [9]. Several lines of evidence suggest that memory consolidation can occur in REM sleep [16]. Theoretical considerations do not preclude retention of memories of the patterns generated in Phase 2 (if it occurs), as they did for Phase 1. However, the bizarre and incomplete quality of these experiences must not be confused with reality.

If evolutionary pressures have led to implementation of an algorithm such as is proposed here for the diminution

of confusion in long-term memory, then it is hard to see how they would have failed to take advantage of the conditions during sleep for its implementation. If the procedures do take place during sleep, then some of the most obvious matters for experimental verification are firstly that interference with sleep behaviour and cueing of particular topics so as to influence sleep processing might affect performance in tasks where confusion is a prominent feature and secondly that there should be some after-effect of SWS extending through much of the subsequent REM sleep, whereby cells most highly activated in the former tend to be harder to excite in the latter.

7. Could selective consolidation aid discrimination learning as well as recall?

In recent work we have addressed the issue whether selective consolidation (using recalled patterns that are rich in distinctive features) could aid the learning of discriminations, as well as long-term recall. In particular, we are interested whether it could be of benefit in situations where normally a large number of repetitions of training sessions is required to produce satisfactory performance. The new task is to generate, for each input pattern, a correct output pattern that has been associated with it in training. Algorithms for solving this problem usually involve learning rules, for changing synaptic weights, that take account of errors made during training [21,22].

The use of recall instead of real training sessions to provide consolidation in the learning of discriminations seems risky. It is different from the situation with episodic memory, where there is initially good transient memory that fades away. With discrimination learning, performance is initially poor and gradually improves with training. Before this process is complete, internally generated recall of training conditions will be prone to errors: if used for consolidation it may simply perpetuate these errors. We have considered, however, whether in situations where there is overlap between input patterns, enrichment of recalled patterns to emphasise distinct features might offer compensating advantages. The first question to ask is whether it would be an advantage if real training sessions were carried out with masked patterns, eliminating the common cells in overlapping input patterns.

We have shown in simulations that masking input patterns to leave just distinct cells active can improve the speed with which a simple learning system can discriminate between stimuli and avoid confusion errors (Fig. 4). The task was to generate correct output patterns (20 active out of 100 output cells) for 100 different input patterns (also 20 active out of 100 input cells, for full patterns). The required output patterns were all randomly related and

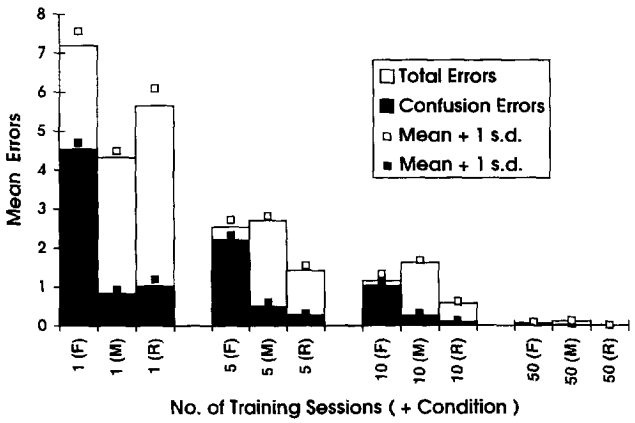

Fig. 4. Average numbers of incorrectly activated cells for all 100 output patterns in a paired association task learned in a network with a single layer of input to output connections, following different numbers of training sessions. Learning employed the delta rule as described in the text, with $\eta = 0.02$. The mean result for five runs with different data sets are shown, with the mean + 1 S.D. for these five runs. Three different conditions were used, corresponding to full input patterns (F), masked patterns (M) and restored patterns (R) described in the text.

thus overlapped on average by 20%. 50 of the input patterns were randomly related: the others were each twinned with one of the first 50 patterns, sharing 15 cells in common ($\beta = 0.75$), with the rest selected at random from those not active in the twin. Thus the problem was to learn 100 paired associations, with much scope for confusion, since each input pattern overlapped substantially with its twin in the training set. The errors in the output are classed as 'confusion errors' when incorrectly activated cells were amongst those that would have been correct for the twinned input pattern.

Learning took place while the paired input and output patterns were presented together during training sessions. Modification of the synaptic weight (w_{ij}) of each direct connection from an input cell (i) onto an output cell (j) was made according to the perceptron or delta rule [21,22]. The changes depend on the binary presynaptic activity (a_i), desired binary output activity (a_j) and synaptic activation s_j onto cell j (i.e., the summed synaptic weights from all active inputs: $s_j = \Sigma w_{ij} a_i$):

$$\Delta w_{ij} = \eta \cdot (a_j - s_j) \cdot a_i$$

The parameter η, affecting speed and stability of learning, was set to 0.02. For purposes of testing, the output patterns were always taken as the correct number of output cells (20): those with the greatest summed activation (s_j) on presentation of an input pattern. This is the output pattern that would be active with recurrent inhibitory feedback set to regulate the correct activity ratio at the output. In Fig. 4 it is shown that with full patterns presented as described above (condition F), errors in the task were almost completely eliminated after 50 training sessions.

After the first few sessions, nearly all the errors were confusion errors involving outputs that would have been appropriate for the twinned pattern.

We ran the simulations with the same input patterns masked to eliminate overlap (condition M), leaving only the five active cells that were distinct from each twinned pattern. The required output patterns were unchanged. The effect was to change the overall numbers of errors rather little: sometimes higher and sometimes lower than with the full patterns. However, the confusion errors were substantially fewer (Fig. 4: black bars). The fraction of total errors that were confusion errors dropped to around 20% (i.e., the activity ratio at the output), since masked patterns bear much the same relationship to their twins as they do to other input patterns. With masked patterns, there was a big increase of ordinary errors, i.e., those not associated with the twin. This is due to the loss of cells that help to discriminate each full input pattern from the non-twinned input patterns. This increase in distributed errors just about balanced the drop in confusion errors involving the twin.

The confusion errors in this simulation amount to overlap with a single other output pattern, appropriate in other circumstances. The distributed errors do not amount to overlap with any single output pattern. On average, 20% of total errors overlap with any output pattern that is not associated with the twinned input: less than or equal to the observed numbers of confusion errors. Thus the effect of masking on the maximum erroneous overlap between outputs during training is represented in Fig. 4 by the confusion errors (black bars) rather than by the total errors. From the point of view of overlap and confusion fed through to later stages of neural analysis, there was substantial improvement with masked patterns.

Since the masked patterns had a lower activity ratio than the full patterns, we ran a control simulation (R in Fig. 4) in which the activity ratio of the masked patterns was restored to normal with the addition of random cells to each pattern. This confirmed that the principal effects of masking were due to the elimination of common cells, not to the reduction of activity ratio.

8. What kinds of learning could benefit from enriched recall during sleep?

Many complex issues are raised by the considerations and simulations set out here. The only solid conclusions are that automatic algorithms can be devised to use recall to help diminish confusion problems in memory in at least some circumstances and that some of these algorithms require special conditions for their implementation. These conditions seem to be inconsistent with use of the relevant parts of the nervous system for interactions with the environment and it is therefore suggested that if they have evolved they would very likely have evolved to occur during sleep. The core idea is that it is sometimes useful to generate enriched recall of experienced patterns, in which the features distinct from other patterns are emphasised. Use of these enriched and abnormal patterns can benefit subsequent waking performance.

The generation of these enriched patterns requires, at least with the algorithms we can devise, conditions in which the normal ability of the nervous system to store traces or memories of its activity patterns is temporarily suspended or curtailed. This state we tentatively identify as occurring during slow-wave sleep. The enriched patterns are generated following this state, at a time when transient after-effects render shared cells or features less susceptible to recall. Consolidation occurs during this phase of the procedure, which we tentatively identify with REM sleep.

The enriched patterns could play a role in the selective consolidation of long-term memories so that confusion between overlapping patterns will be reduced at times when transient memories have faded. This can in principle reduce subsequent confusion in two situations: firstly, when pairs of overlapping memories have been freshly experienced and are being consolidated and secondly when a single new experience overlaps with an old and already consolidated experience [11].

Recall of enriched patterns could also play a role in improving discrimination learning. Learning to discriminate between different input patterns, producing different appropriate outputs for each, can require many repetitions. This is especially the case when overlap between input patterns leads to a substantial number of confusion errors. We have shown that the use of masked input patterns (eliminating overlap between overlapping pairs of input patterns) can speed training to criteria based on output overlap. We have not yet set up full simulations to assess how much such benefit would extrapolate to savings in training trials with full patterns. It does, however, look at least plausible in principle that selectively enriched consolidation could benefit discrimination learning as well as long-term recall.

References

[1] Baddeley, A., *Human Memory: Theory and Practice*, Lawrence Erlbaum, Hove, 1990, 515 pp.

[2] Bramham, C.R. and Srebro, B., Synaptic plasticity in the hippocampus is modulated by the behavioral state, *Brain Res.*, 493 (1989) 74–86

[3] Clark, J.W., Rafelski, J. and Winston, J.V., Brain without mind: computer simulation of neural networks with modifiable neuronal interactions, *Phys. Rep.*, 123 (1985) 215–273

[4] Crick, F. and Mitchison, G., The function of dream sleep, *Nature*, 304 (1983) 111–114
[5] Crick, F. and Mitchison, G., REM Sleep and Neural Nets, *J. Mind Behav.*, 7 (1986) 229–250
[6] Evarts, E.V., Temporal patterns of discharge of pyramidal tract neurons during sleep and waking in the monkey, *J. Neurophysiol.*, 27 (1964) 152–171
[7] Gaffan, D., Recognition memory in animals. In J. Brown (Ed.), *Recognition and Recall*, Wiley, London, 1975
[8] Gardner-Medwin, A.R., Modifiable synapses necessary for learning, *Nature (London)*, 223 (1969) 916–918
[9] Gardner-Medwin, A.R., Optic radiation activity during sleep and waking, *Exp. Neurol.*, 43 (1974) 314–329
[10] Gardner-Medwin, A.R., The recall of events through the learning of associations between their parts, *Proc. R. Soc. London B*, 194 (1976) 375–402
[11] Gardner-Medwin, A.R., Doubly modifiable synapses: a model of short and long term auto-associative memory, *Proc. R. Soc. London B*, 238 (1989) 137–154
[12] Gardner-Medwin, A.R., Possible strategies for using sleep to improve episodic memory in the face of overlap. In J.G. Taylor and C.L.T. Mannion (Eds.), *Theory and Applications of Neural Networks*, Springer-Verlag, London, 1992, pp. 129–138
[13] Gibson, W.G. and Robinson, J., Statistical analysis of the dynamics of a sparse associative memory, *Neural Networks*, 5 (1992) 645–661.
[14] Giuditta, A., A sequential hypothesis for the function of sleep. In W.P. Koella, E. Ruther and H. Schultz (Eds.), *Sleep it '84*, Fischer Verlag, Stuttgart, 1985, pp. 222–224.
[15] Hebb, D.O., *The Organization of Behaviour*, Wiley, New York, 1949
[16] Hennevin, E., Hars, B., Maho, C. and Bloch, V., Processing of learned information in paradoxical sleep: relevance for memory, *Behav. Brain Res.* 69 (1995)
[17] Lansner, A., Ekeberg, O., Reliability and speed of recall in an associative network, *IEEE Trans. PAMI*, 7 (1985) 490–498
[18] Leonard, B.J. McNaughton, B.L. and Barnes, C.A., Suppression of hippocampal synaptic plasticity during slow-wave sleep, *Brain Res.*, 425 (1987) 174–177
[19] Maho, C. and Bloch, V., Responses of hippocampal cells can be conditioned during paradoxical sleep, *Brain Res.*, 581 (1992) 115–122
[20] Marr, D., A theory for cerebral neocortex, *Proc. R. Soc. London B*, 176 (1970) 161–234
[21] Rosenblatt, F., *Principles of Neurodynamics*, Spartan, Washington, 1962.
[22] Rumelhart, D.E., Hinton, G.E. and McLelland, J.L., A general framework for Parallel Distributed Processing. In D.E.R. Rumelhart et al., *Parallel Distributed Processing, Vol. I*, MIT Press, Cambridge, MA, 1986, pp 45–76.
[23] Stones, M.J., Memory performance after arousal from different sleep stages, *Br. J. Psychol.*, 68 (1977) 177–181
[24] Treves, A. and Rolls, E., What determines the capacity of auto-associative memories in the brain, *Network Computation Neural Systems*, 2 (1991) 371–397