An Assessment of Marr's Theory of the Hippocampus as a Temporary Memory Store

D. J. Willshaw; J. T. Buckingham

Stable URL:

http://links.jstor.org/sici?sici=0962-8436%2819900829%29329%3A1253%3C205%3AAAOMTO%3E2.0.CO%3B2-0

Philosophical Transactions: Biological Sciences is currently published by The Royal Society.

An assessment of Marr's theory of the hippocampus as a temporary memory store

D. J. WILLSHAW AND J. T. BUCKINGHAM

Centre for Cognitive Science, University of Edinburgh, 2 Buccleuch Place, Edinburgh EH8 9LW, U.K.

SUMMARY

The recent reawakened interest in 'neural' networks begs the question of their relevance to the analysis of real nervous systems. Network models have been criticized for the lack of realism of their individual components, and because the architectures required by some neural-network algorithms do not seem to exist in real nervous systems. In three related papers published in the 1970s, David Marr proposed that the cerebellum, the neocortex and the hippocampus each acts as a memorizing device. These theories were intended to satisfy the biological constraints, but in computational terms they are underdetermined. In this paper we reassess Marr's theory of the hippocampus as a temporary memory store. We give a complete computational account of the theory and we show that Marr's computational arguments do not sufficiently constrain his choice of model. We discuss Marr's specific model of temporary memory with reference to the neurophysiology and neuroanatomy of the mammalian hippocampus. Our analysis is supported by simulation studies done on various memory models built according to the principles advocated by Marr.

1. INTRODUCTION

In the study of neural networks, there have been very few attempts to examine in detail how a particular part of the brain might compute a particular function. This paper examines one theory of learning and memory that does consider seriously the constraints imposed by the available computing machinery of the brain. This is the proposal, due to the late David Marr (1971), that the mammalian hippocampus acts as a temporary content-addressable memory store.

Like its companion papers on the cerebellum (Marr 1969) and the neocortex (Marr 1970), this paper remains a potential source of inspiration for those interested in the theory of the nervous system. However, even almost 20 years after its publication in *Philosophical Transactions*, it is far more widely cited than understood. In view of the growing interest in the neural-network paradigm of computation, we decided to investigate and evaluate the claims made in this paper. Marr's main tools of investigation were mathematical analysis and numerical solution of the equations he formulated for the various computations envisaged. It is now possible to add the method of computer simulation, which in the early 1970s was not feasible for networks of any appreciable size.

2. THEORIES AND MODELS

The title of Marr's paper is 'Simple memory: a theory for archicortex' (hippocampus), and in the body of the paper he discusses a model of mammalian hippocampus. There are different kinds of model, depending to a large extent on the use to which they are put. Here it is useful to distinguish between simplifying and realistic models. Simplifying models are generally used as tools to discover the principles of operation of systems, and so most of the details are abstracted away. Realistic models include most of the parameters known about the systems at the level of organization used in the models. They are generally used to explore parameter sensitivity (Sejnowski *et al.* 1988). The criticism has often been made that the simplifying neurobiological models investigated in neural network research are too abstract to give any valid insights about the brain. Implicit in this criticism is that realistic models would be more useful. However, as Sejnowski *et al.* (1988) noted, as we do not have such a complete understanding of the brain, a realistic model would contain so many free parameters that no specific predictions would be generated. There is also the danger that any completely specified model will become just as complex and as difficult to analyse as the brain itself. These considerations limit the usefulness of realistic models. With the appropriate assumptions, simplifying models can be powerful explanatory tools. Marr's model is of this type.

3. MARR'S COMPUTATIONAL THEORY OF THE HIPPOCAMPUS: CONSTRAINTS

The hippocampus is supposed to act as a temporary content-addressable memory. A number of *events*, each represented as a pattern of activity in a selected population of nerve cells, is to be stored. Subsequent presentation of a small part of a previously stored event must then enable the whole of it to be reconstructed, in terms of the activity in these cells.

The basic conceptual model used by Marr combines features of associative learning (Willshaw *et al.* 1969; Kohonen 1972; Gardner-Medwin 1976) and competitive learning (Grossberg 1973). Each event is represented by a particular pattern of activity in the fibres of the input cells, A (figure 1 *a*). Presentation of each event in turn causes a mapping to be built up onto a second set of cells, B, onto which the A-cells project, by modification of the synapses between them. There may also be intermediate layers of cells. The B-cells have a return projection to contact the A-cells, also through modifiable synapses. In the retrieval process, presentation of part of a previously stored event can than be enough to activate sufficient of the B-cells that originally responded in the storage of the full event to re-create the original complete pattern of firing over the A-cell population. The B-cells effectively come to respond to patterns of activity corresponding to parts of the stored patterns.

He first discusses why a structure specialized for temporary storage is needed. It is regarded as an ancillary unit to the neocortex, and is responsible for the instantaneous storage of new data as they arise. In his earlier theory of the neocortex (Marr 1970), he

(a)

(b)

Figure 1. Architecture of the models discussed. (*a*) The basic model, in which neocortical pyramidal cells project, possibly through intermediate layers of cells, to cells in the temporary memory structure. These cells then project back to the neocortex. (*b*) The three-layer model analysed by Marr.

proposed that the function of the neocortical pyramidal cells is to reorganize and classify this information, incorporating it or discarding it according to its usefulness to the animal. In the hippocampus paper he argues that it would be inefficient to store such transient information in the permanent memory of the neocortex until it is known what features of the new information are required. In addition, neocortical connectivity may not be sufficiently extensive to allow associations between any two pyramidal cells to be built up.

The central question posed by Marr concerns what sort of structure is required for this temporary memory, and whether it matches the known structure of the hippocampus.

Several assumptions and computational relations constrain the form of the solution being sought. Broadly speaking, these can be classified as numerical constraints, which were derived intuitively or have some biological basis, and computational constraints, which must hold if the memory is to function satisfactorily.

3.1. *Numerical constraints*

N1: *the number of events to be stored*

Based on the idea that information is transferred to long-term memory during sleep (Marr 1970) and that no more than one event per second is stored, it is assumed that 10^5 events are to be stored. This number is chosen as it is roughly equal to the number of seconds in a day.

N2: *the number of inputs to the memory*

The events are assumed to be represented by patterns of activity in the neocortex. Marr observes that if all the neocortical pyramidal cells sent information to the short-term memory then this would be an enormous structure. It is assumed that there are indicator cells amongst the neocortical population that each sample the activity in $\frac{1}{30}$ mm^2 of neocortex. These cells constitute the input cells of the memory. Assuming that there is 4×10^4 mm^2 of neocortex (containing 10^8 neocortical cells) requiring access to the memory, this gives roughly 10^6 indicator cells.

N3: *the number of outputs from the memory*

Since neocortical pyramidal cells each have fewer than 10^5 synapses (Cragg 1967), most of which are held to be concerned with diagnosis and classification, it is estimated that only about 10^4 synapses are available for receiving the output from simple memory; i.e. there are at most 10^4 B-cells if each has to contact each A-cell in the return projection. Later in Marr's paper, this number is changed to 10^5.

N4: *synaptic modification*

It is assumed that this is an all-or-none process; i.e. synapses are either 'on' or 'off'. The informal justification given is that continuously valued synapses are not required in the hippocampal model as classification, which involves the computation of conditional probabilities which can be stored in temporary valued synapses, is not performed there. The rule for synaptic modification is a simple version of
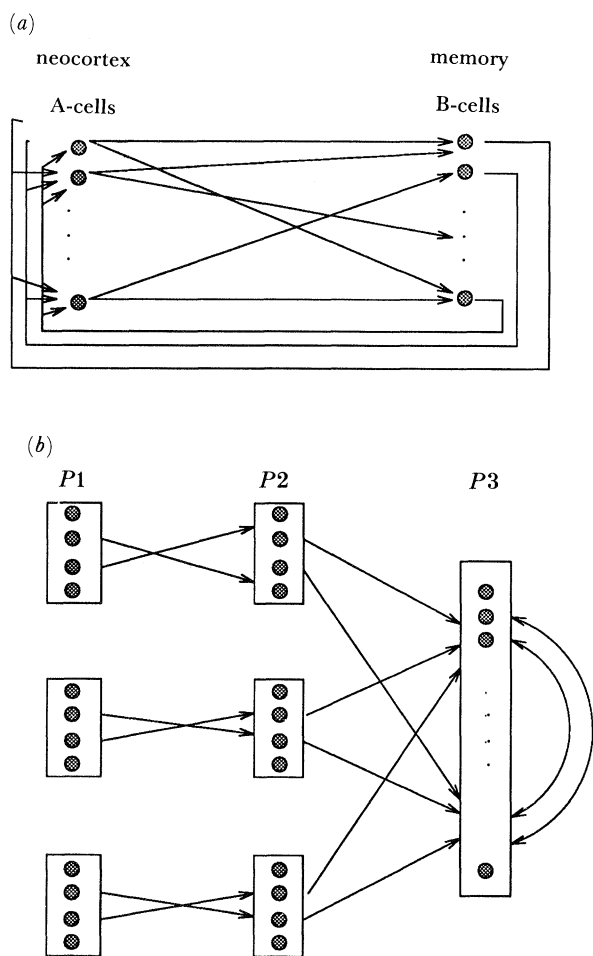
what is now termed a Hebb (1949) rule (see also Brindley (1969)): synapses are switched from the 'off' to the 'on' state when the appropriate presynaptic and postsynaptic cells are both active.

N5: *the number of synapses made on a given cell*

In accordance with biological information (Cragg 1967), this is assumed to be high, and an upper limit of 10^5 synapses per cell is set.

N6: *the pattern of connectivity*

To impose a rough topographic ordering on the projections (Raisman et al. 1965), some of the layers are subdivided into blocks, and connections between layers are only allowed between the cells in the corresponding blocks. Within this restriction, connections are made at random.

N7: *the level of activity*

It is assumed that in each layer all full events cause the same amount of activity. To keep activity levels constant, there must be a means of adjusting the thresholds at which cells fire. This requires the activity in populations of cells to be sampled, and for accurate sampling the level of activity cannot be too low. It is assumed that no fewer than 1 cell per thousand is active in any population of cells.

N8: *size of cue*

The fraction of a previously stored event required to successfully retrieve the full event from store is set at $\frac{1}{10}$. As Marr recognizes, this figure is somewhat arbitrary. If the minimum acceptable size of cue were increased, then more events could be stored and retrieved satisfactorily, and vice versa.

3.2. *Computational constraints*

For a network made up of several layers of cells $\mathscr{P}_1$, $\mathscr{P}_2$, $\mathscr{P}_3$..., each layer $\mathscr{P}_i$ receives connections from layer $\mathscr{P}_{i-1}$ and sends connections to layer $\mathscr{P}_{i+1}$. For layer $\mathscr{P}_i$ the key numerical constants are: N_i, the number of cells; α_i, the proportion of cells that are active at any given time (the level of activity); S_i, the number of synapses per cell from cells in layer $\mathscr{P}_{i-1}$.

Various other parameters are used in the paper, whose values are determined by these three basic parameters. Principally, these are: L_i, the number of cells that are active; Z_i, the probability of connection from a cell in layer $\mathscr{P}_{i-1}$ to layer $\mathscr{P}_i$; Π_i, the probability that a given synapse on a cell in $\mathscr{P}_i$ will have been facilitated.

The relations between these parameters are:

$$L_i = \alpha_i N_i,$$
$$Z_i = S_i/N_{i-1},$$
$$\Pi_i \approx 1 - \exp\left(-n\alpha_{i-1}\alpha_i\right).$$

The expression for Π_i is derived from the probability that a synapse is never modified in the storage of n events, which is $(1-\alpha_{i-1}\alpha_i)^n$, or approximately $\exp\left(-n\alpha_{i-1}\alpha_i\right)$.

C1: *number of modified synapses*

To prevent cells that should remain inactive from becoming erroneously active during recall, the fraction of modified synapses should be not too large. A rough condition for good recall is that the exponent in the expression for Π_i be no greater than 1:

$$n\alpha_{i-1}\alpha_i \leqslant 1.$$

Under this condition, each synapse is associated with the storage of an average of no more than one event.

C2: *full representation of the input*

It must be ensured that no information is lost on transfer of a pattern of activity from layer $\mathscr{P}_{i-1}$ to layer $\mathscr{P}_i$. A necessary condition is that the probability P that any active cell in $\mathscr{P}_{i-1}$ does not synapse with an active cell in $\mathscr{P}_i$ is small:

$$P \approx \exp\left(-\alpha_i S_i N_i/N_{i-1}\right).$$

or

$$P = (1-Z_i)^{L_i},$$

Setting P to be less than the small number $\exp(-20)$ yields the inequality:

$$S_i\alpha_i N_i \geqslant 20 N_{i-1}$$

4. MARR'S FIRST MODEL

This consists of a layer of input cells from the neocortex (A-cells), which contact the output cells (B-cells) through modifiable synapses. The B-cells themselves project directly back onto the pyramidal cells of the neocortex (figure 1a). There are no intermediate layers of cells.

Marr assesses the plausibility of this model by calculating the number of output cells that would be active per event. In his work on the cerebellum (Marr 1969), he calculated that a single output cell can respond to roughly 10^2 stored events without appreciable error. This hippocampal model has 10^4 output cells (constraint N3), which will therefore be active on 10^6 occasions. As 10^5 events are to be stored (constraint N1), just ten cells would be active per event. This number is assessed to be too small to allow a reliable representation of the input event, and this model is rejected. The reason for rejection is not spelt out, but presumably Marr means that having 10 out of 10^4 output cells active would not allow a sufficiently accurate sampling of output fibre activity that is required for the purposes of threshold setting. As we explain later, if this model had been given 10^5 output fibres, the number that was chosen for the improved model, this argument would not have applied.

5. THE BASIC COMPUTATIONAL MODEL

Marr proposes to solve the problems encountered with his first model by augmenting it in two ways (figure 1b).

1. A layer of cells intermediate between the A-cells and the B-cells is provided. The justification, which is not given, is probably that this would provide an extra

population of modifiable synapses, and thus potentially more bits of storage.

2. To reduce the computations that have to be done in the return projection to the neocortex, it is assumed that completion of any sub-event during recall is performed in the short term memory, and thus is finished before the representation of the event, the *simple representation*, leaves the short term memory for the neocortex. Completion is mediated by the collateral effect: there are recurrent connections between B-cells, to make a form of auto-associative network (Kohonen 1972).

This is the model, the *simple memory*, that Marr analyses in detail. The task is now to specify the parameter values.

In the neocortical theory (Marr 1970) it was assumed that each event was assigned according to a diagnostic procedure to some members of a family of classes, each class being represented by the firing of a neocortical pyramidal cell. In the case of simple memory, information has to be stored as it occurs and without attempting the best classification of it. It is proposed that the sub-family of output cells of the memory to which each incoming event is assigned is merely those that have more relevant synapses than others.

The analysis given by Marr is concerned with layers of cells $\mathscr{P}_1$, $\mathscr{P}_2$ and $\mathscr{P}_3$, which interconnect with certain contact probabilities. The return projection is not considered. $\mathscr{P}_1$ connects to $\mathscr{P}_2$, and $\mathscr{P}_2$ to $\mathscr{P}_3$. To approximate the biology, the projection of $\mathscr{P}_1$ to $\mathscr{P}_2$ is assumed to have a block structure (constraint N6). In each of the 25 identical blocks making up $\mathscr{P}_1$, the cells project exclusively to the cells in the corresponding block in $\mathscr{P}_2$. Each block of $\mathscr{P}_2$ projects to the whole of $\mathscr{P}_3$. The model is specified by the values of its parameters, which are set by the constraints given.

Our explanation for how the parameter values are derived is shorter than Marr's, and does not use precisely the same arguments. As far as we can judge, it is essentially the same.

5.1. *Values for the activity ratio,* α_i

Although the activity ratio has slightly different values from layer to layer (for which arguments are given in this paper), its value of around 0.003 is determined essentially by constraints C1 (accurate recall) and N7 (activity level as high as possible).

This value is most easily derived in the case of the collateral effect in $\mathscr{P}_3$.

Constraint C1 is

$$n\alpha_{i-1}\alpha_i \leqslant 1.$$

For the collateral effect, $\alpha_{i-1} = \alpha_i = \alpha_3$, and with the number of events, n, at 10^5, this gives $\alpha_3^2 \leqslant 10^{-5}$, or

$$\alpha_3 \leqslant 0.003.$$

5.2. *Parameters for layer* $\mathscr{P}_3$

The value of α_3 is set at around 0.002. The number of cells, N_3, now needs to be reasonably high, to give a

realistically large number of active cells. It is set at 100 000, giving 200 cells active per event. In accordance with constraint N5, S_3 is assumed to be reasonably large, and is chosen to be 50 000.

The parameters for layer $\mathscr{P}_3$ are therefore: $N_3 = 100\,000$; $S_3 = 50\,000$; $\alpha_3 = 0.002$.

5.3. *Parameters for layer* $\mathscr{P}_2$

Because the values of S_3, α_3 and N_3 are known, constraint C2 (full representation) can be used to obtain directly an upper bound of 500 000 on the number of cells in $\mathscr{P}_2$:

$$S_3 \alpha_3 N_3 \geqslant 20 N_2.$$

N_2 is set at this number, giving 20 000 cells in each of the 25 blocks making up $\mathscr{P}_2$. For the value of the activity α_2 in this layer, Marr uses a value of 0.006. The other parameter is S_2, the number of synapses per cell. This is set arbitrarily at 10 000, bearing in mind that each cell in $\mathscr{P}_2$ can receive connections from only one block of $\mathscr{P}_1$.

The parameters for layer $\mathscr{P}_2$ are therefore: $N_2 = 500\,000$, giving 20 000 cells per block; $S_2 = 10\,000$; $\alpha_2 = 0.006$.

5.4. *Parameters for layer* $\mathscr{P}_1$

The value of α_1 is set at 0.002. The value of the number of cells, N_1, has already been set by constraint N2 to be about 10^6. A value of 1.25×10^6 is chosen, giving 50 000 cells per block, which satisfies constraint C2 (full representation). A value for S_1 is not needed.

The parameters for layer $\mathscr{P}_1$ are: $N_1 = 1.25 \times 10^6$, giving 50 000 cells per block; $\alpha_1 = 0.002$.

The only structural parameter whose value remains to be determined is the collateral contact probability. Marr considers this in his discussion of how the thresholds should be set in the various layers for accurate completion of sub-events.

5.5. *Storage and retrieval of events: setting the threshold*

Activity in the cells in one layer influences the cells in the next layer, through the synapses between them. During storage, the cells that fire in this next layer are those that have the most activated synapses (whether or not they have been previously modified), the threshold for firing being set to give the level of activity required for the layer as a whole. All synapses where the presynaptic and the post-synaptic cells are both active are then set to the 'on' state.

The threshold setting procedure used during retrieval is more complicated and relies on information about the number of activated synapses *and* the number of modified synapses impinging on each cell. Marr first considers the retrieval process in the output layer $\mathscr{P}_3$. Suppose that presentation of part of a previously learnt event has caused certain cells in $\mathscr{P}_3$ to fire, some of which are part of the representation of that event, and others not. Through the recurrent connections within

$\mathscr{P}_3$, the activity in these cells is fed back, which may cause some quiescent cells in $\mathscr{P}_3$ to start firing and some others that are firing to stop. If the correct representation of the whole event is to emerge, then the number C_0 of 'genuine' cells firing should increase and the number C_1 of 'spurious' cells firing should decrease.

With L_3 and $Z_{3'}$ denoting the number of active cells in $\mathscr{P}_3$ and the collateral contact probability, respectively, the number of active collateral synapses to any given cell in $\mathscr{P}_3$ is drawn from a population of size L_3 and binomially distributed with expectation $L_3 Z'_{3'}$. For a 'spurious' cell with x synapses, the number of these that had been modified during storage is binomially distributed with expectation $x\Pi_{3'}$, where $\Pi_{3'}$ is the modification probability for the collateral synapses. If the cell is 'genuine', all the collateral synapses from other genuine cells will have been modified during storage, and the other synapses will each have been modified with a probability as given for the spurious cells.

The rule for deciding whether a cell in $\mathscr{P}_3$ fires is that a sufficient proportion f of its active synapses were modified *and* the total number of its active modified synapses exceeds a certain absolute level, T. The division threshold for the cell is denoted by f and T denotes the subtraction threshold. To set these thresholds, Marr assumes that the cell can measure the number of currently active modified synapses, and
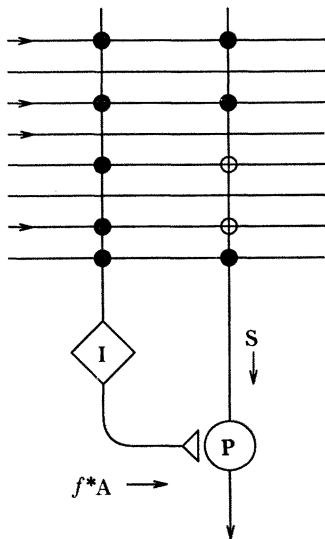

Figure 2. Threshold setting. A pyramidal unit (P) and an inhibitory unit (I), which implements a division threshold. Their dendrites are represented by the vertical lines, and a set of axons that project onto them are represented by the horizontal lines. Synapses are represented by small circles. Empty circles represent unmodified synapses, filled circles modified ones. The activated axons are indicated by the arrows. The amount of excitation reaching the pyramidal cell body is $S = \Sigma w_i x_i$ where w_i is the strength of the synapse from axon i onto the dendrite and x_i is the state, (either 0 or 1), of axon i. The inhibitory unit I measures the activity, $A = \Sigma x_i$ summed only over the axons that synapse onto cell P, and sends a fraction f of this as inhibition to the pyramidal cell. To fire, the excitation, S, must exceed some absolute threshold T and also exceed fA. In this example, $S = 2$ and $A = 3$ (as axons that do not synapse onto P do not count); P will fire for $T \leqslant 2$ and $f \leqslant \frac{2}{3}$.

another supporting cell measures the number of currently active synapses (figure 2).

During one pass through the collateral system, any particular settings of these dual thresholds will cause changes in the numbers C_0 and C_1 of genuine and spurious cells. The dual thresholds must be set to maintain the total number of cells firing at the required level.

For a stored event to be reconstructed satisfactorily, the ratio of genuine to spurious cells firing must increase in the feedback process, ultimately resulting in a perfectly recalled pattern. Marr tabulates the effects on the ratio C_0/C_1 of changing T and f, for various initial values of C_0/C_1. By this means he attempts to find the smallest value of C_0/C_1 from which the complete pattern can be retrieved through the collateral effect. This value is called the *statistical threshold*.

For a network with $N_3 = 10^5$ cells and $L_3 = 200$ cells active per event, he obtains numerical values for the best settings of the two thresholds T and f, for various different numbers of events stored, n, and values of the collateral contact probability, $Z_{3'}$. 'Suitable juggling' of T and f is required: f must start low, and increase gradually as C_0/C_1, the fraction of genuine to spurious cells, increases. At the same time, to keep the activity in $\mathscr{P}_3$ constant, T must decrease, as increasing f on its own will reduce the number of active cells. Provided that initially about 10% of the genuine cells are included in the initial activity pattern, recovery of the whole event is ensured, and in about 3 cycles. He does not experiment with different densities of collateral connections, but observes that a contact probability of 0.2 allows a statistical threshold which is roughly half that needed for a probability of 0.1.

Marr makes similar calculations for recall performance from $\mathscr{P}_1$ to $\mathscr{P}_2$ and $\mathscr{P}_2$ to $\mathscr{P}_3$, with the difference that recall for these stages is a feed-forward process, with no feedback. The threshold setting mechanisms are assumed to operate as for the collateral projection in $\mathscr{P}_3$.

5.6. Summary description of the memory

Constraints

(i) The memory should consist of layers of cells, each receiving connections from one layer and projecting to the next.

(ii) There should be 10^6 input fibres and 10^5 (or 10^4) output fibres.

(iii) The memory needs a capacity of the order of 10^5 events.

(iv) There must be good content-addressable recall capabilities.

(v) Recall should be complete before information has left the memory for the neocortex.

Specification

The memory has 1.25×10^6 input fibres, divided into 25 blocks of 50000 fibres each. It has an intermediate layer of 500000 cells, also divided into 25 blocks and an output layer of 100000 cells, in a single block.

Performance

The result of Marr's numerical calculations is as follows. After 50000 events have been stored (each represented by activity in 100 of the input cells in each block), an individual event will be recalled successfully if there is activity in just 30 of the fibres originally active in that event, provided they belong to one block. After 100000 events have been learnt, 60 such 'genuine' fibres are required for accurate recall. When 'spurious' input fibres are included, the numbers are different: for 50000 events stored, out of 100 active input cells, all in the same block, 70 of these must be genuine; for 100000 events, there must be 90 genuine cells out of the 100 cells active.

What Marr claims to have shown is that a model consisting of one layer of cells ($\mathscr{P}_1 \to \mathscr{P}_2 \to$ return) cannot satisfy the constraints, but a memory with an intermediate layer ($\mathscr{P}_1 \to \mathscr{P}_2 \to \mathscr{P}_3 \to$ return) can. He then goes on to argue that the structure of the three-layer model, which was derived from computational considerations, indeed matches that of the hippocampal formation.

6. AN INTERPRETATION OF THE HIPPOCAMPAL FORMATION IN TERMS OF THE MODEL OF SIMPLE MEMORY

In §4, Marr gives a most detailed account of the morphology of the mammalian hippocampal formation, based on information supplied by Cajal (1911) and Lorente de No (1933) on the mouse and by Blackstad (1956) and White (1934) on the rat. He uses the terminology of Blackstad to describe the cell types and the nature and extent of their processes.

6.1. *Representation of layers $\mathscr{P}_1$, $\mathscr{P}_2$ and $\mathscr{P}_3$ in hippocampal structures*

The pyramidal cells of the CA areas (figure 3) are regarded as populations of cells in which simple representations of events are formed; i.e. these are the cells of $\mathscr{P}_3$ in the model. If this correspondence is correct, then the hippocampus should have the following properties: the input fibres should be suitable (in terms of their number and origin); the activity in the cells should be low; each cell should have very many (up to 50000) modifiable synapses from the previous layers of cells; there should be an extensive collateral system, with each cell having around 10000 modifiable synapses from other CA cells; there should exist appropriate supporting cells to supply inhibition for threshold setting. Marr discusses in some detail to what extent the hippocampus has these properties. In particular, he discusses what computations are required for obtaining the correct conditions for synaptic modification and for setting the thresholds on the CA pyramidal cells. Briefly, two kinds of inhibitory signal are required: one to the dendrites of pyramidal cells to mediate the subtractive threshold, and one to the soma to mediate the division threshold. Marr describes types of interneuron that could undertake
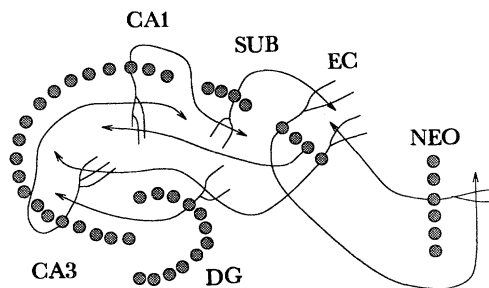

Figure 3. Schematic diagram of the major cortical pathways through the hippocampal formation (Squire *et al.* 1989). Neocortical pyramidal cells (NEO) project (via the perirhinal cortex) to layer II and III entorhinal cortex cells (EC). The axons of the layer II EC cells form the perforant path that penetrates the hippocampal fissure and projects to dentate gyrus (DG) granule cells and CA3 pyramidal cells; EC layer III cells project directly to CA1 pyramidals. The DG granule cell axons, the mossy fibres, form strong excitatory synapses on the dendrites of the CA3 pyramidals. The axons of the CA3 pyramidals split into three pathways: (i) collateral projections synapse widely onto other CA3 cells, (ii) one pathway leaves the hippocampal formation via the fornix, and (iii) a final pathway (shown here), the Schaffer collaterals, projects to CA1. The CA1 pyramidal cells project primarily to the subiculum (SUB) although some project directly back to EC. The subiculum (and CA1) project to the deep cells of EC, which project back to the neocortical areas that originally projected to EC, thus completing the loop.

these functions, but he does not describe an explicit model for the entire process.

The granule cells of the dentate gyrus are regarded as cells where simple representations are also set up that are then sent to the CA region.

The entorhinal cortex (figure 3) and presubiculum are regarded as those regions that prepare information from many different sources for its simple representation in the CA system and the dentate gyrus. Thus entorhinal cortex and presubiculum are regarded as a rough model for layer $\mathscr{P}_2$. The cells in these two regions are described and roles assigned to them.

The neurobiological representation of $\mathscr{P}_1$ is not discussed. Because it is assumed that the cells of this layer are found in the neocortex, consideration of these cells may not have been thought appropriate in a discussion of the hippocampus. A subsequent paper on hippocampal input–output relations was promised, but never published.

Marr concludes with an extensive list of neurophysiological predictions from the theory, principally of the function of the memory cells of layers $\mathscr{P}_2$ and $\mathscr{P}_3$ and of the supporting cells and the nature of the synapses of these cells.

7. ASSESSMENT OF MARR'S THEORY

This paper is the third in a sequence of theoretical papers on how specific brain structures can be used as memorizing devices: on the cerebellum (Marr 1969), the neocortex (Marr 1970) and the hippocampus (Marr 1971). A fourth neurobiological paper, on the retina, was published a few years later (1974), but it does not properly belong in this sequence. All three

papers used the same computational ideas, so much so that in each paper he draws on the theorems developed and the results derived in the other two. The same basic approach was used in these papers. This was to show how the neurobiological constraints, mainly neuroanatomical, are consistent with the idea of the chosen part of the brain computing a particular function. In the main, the methods of analysis involved assigning values to the parameters of the system that were consistent with the various neurobiological and computational constraints.

As in the other two papers, the mathematics of this paper is over-elaborate and the desire for verbal and mathematical rigour quite often conceals the points that are being made. Particularly in the light of Marr's subsequent development of the importance of the computational, the algorithmic and the implementational levels of analysis (1982), it is interesting that in this paper on the hippocampus he alternates between two claims: (i) that the structure of the simple memory proposed must necessarily be so for it to act as a content-addressable memory; (ii) that it has to have this structure because the hippocampus is built like this. In the later paper on the retina (1974), the lines between the computational, algorithmic and implementational levels are more clearly drawn.

8. NEUROBIOLOGICAL IMPLICATIONS

Marr's choice of the structure of his model seems to have been influenced heavily by his view that the hippocampus is a three-layer network, a view constructed somewhat independently of his computational results. Although he does characterize the individual properties of the cells that are to form the various proposed layers in some detail, only a loose correspondence between the sub-divisions of the hippocampus and the layers of the model is made. The most extensive discussion revolves around the nature of layer $\mathscr{P}_3$. In identifying the memory elements of this layer with the CA pyramidal cells of the hippocampus, he is placing less importance on the dentate gyrus–CA3-CA1 trisynaptic circuit (figure 3) (Andersen *et al.* 1971) than might have been expected. The granule cells of the dentate gyrus are, however, treated separately, but the pathway involving these cells is not represented in his model.

In these respects, Marr presents a somewhat abstract interpretation of the hippocampus as a temporary memory, in sharp contrast to, say, his theory of the cerebellum (1969). On the other hand, he did discuss what type of local circuitry is required to perform the arithmetic operations, such as division and subtraction, needed for threshold setting. However, this discussion was concerned with principles of computation and no specific models were provided. Perhaps his most important contribution was that he provided many detailed predictions, such as those concerning the level of activity and the modification of synapses, which might have been followed up.

The model requires that synapses are modified by simultaneous pre-synaptic and post-synaptic activity. It pre-dates the finding of long term potentiation (Bliss

& Lømo 1973) in the hippocampus, although he does add a note in proof about Lømo's earlier paper (1971) showing synaptic facilitation in the perforant path–dentate gyrus pathway.

9. COMPUTATIONAL CONSIDERATIONS
9.1. *General points*

The key part of the numerical analysis concerns the collateral effect in $\mathscr{P}_3$. We repeated Marr's numerical calculations for this and obtained similar results. In this respect Marr's model of simple memory does seem to function in the way prescribed. However, several criticisms can be made. The main point is that he is inconsistent in his specification of the model to be analysed. The effect of the extra layer needed to account for the important dentate gyrus–CA3 projection is not considered; nor are the workings of the return projection, which formed part of the original specification. As this projection is to the neocortex, consideration of it might have been thought to be beyond the scope of this paper. A serious inconsistency is that the number of output cells was changed from 10^4 to 10^5, which has a bearing on the choice of the model he ultimately analyses.

9.2. *Rejection of the two layer model*

The first model ($\mathscr{P}_1 \rightarrow \mathscr{P}_2 \rightarrow$ return) was rejected on the grounds that with 10^4 output cells there would be just 10 cells active per event, which is too low for reliable sampling. Here the argument is used that an individual output cell can learn no more than about 100 different events, derived in the cerebellum paper (Marr 1969). The applicability of the cerebellum result to the hippocampus might be questioned since the structure of the two systems (particularly with respect to the important supporting cells) is entirely different. More important is that if he had assumed a figure of 10^5 output cells (as he eventually did for the second model), giving 100 cells active per event, the argument for rejecting the two-layer model could not have been used.

The two-layer model was rejected too speedily. To evaluate the difference between two-layer and three-layer models, we decided to investigate, by the method of computer simulation, models constructed according to Marr's specification.

10. SIMULATION STUDIES
10.1. *Structure of the networks*

Current computer technology does not allow for the simulation of a system of the size analysed by Marr: in a three-layer network with 10^6 cells in $\mathscr{P}_1$, 5×10^5 in $\mathscr{P}_2$ and 10^5 in $\mathscr{P}_3$, there are potentially 10^{12} connections. We found that networks with layers containing a few thousand cells could be simulated satisfactorily on our Sun-3 and Sun-4 machines. The three-layer network we used had 8000 cells in layer $\mathscr{P}_1$, 4000 in $\mathscr{P}_2$ and approximately 1000 in $\mathscr{P}_3$, these numbers being chosen in accordance with Marr's constraints. This is roughly a one-hundredth scale model, but is still

large, with 36 million potential connections. The free parameter at our disposal was the number of events to be stored, n. In line with Marr, and with results from standard competitive networks (Grossberg 1987), we took it to be of the order of the number of cells (in this case, 10^3) in the output layer.

Once the values were specified for the number of events, n, and the number of cells in each layer, N_1, N_2, N_3, the other parameters of the three-layer net were determined by constraints C_1 and C_2:

$$n\alpha_{i-1}\alpha_i \leqslant 1,$$

$$S_i\alpha_i N_i \geqslant 20N_{i-1}.$$

Setting $\alpha_1 = \alpha_2 = \alpha_3$, and using C_1 as an equality determined the values of α_i. Using C_2 as an equality determined the values of S_2 and S_3. The set of basic parameter values used for the three layer network is:

$\alpha_1 = 0.03,$ $\alpha_2 = 0.03,$ $\alpha_3 = 0.03,$
$N_1 = 8000,$ $N_2 = 4000,$ $N_3 = 1024,$
$S_2 = 1333,$ $S_3 = 2666.$

The two-layer network with which this was compared had its parameters determined in an identical fashion. These are:

$\alpha_1 = 0.03,$ $\alpha_3 = 0.03,$
$N_1 = 8000,$ $N_3 = 1024,$
$S_3 = 5333.$

Following Marr, the connections made on each cell were selected at random. However, in our case each cell in a layer received the same number of connections, as specified by the value of S_2 or S_3, respectively. Representation of the block structure was not attempted as it did not seem to be essential to our comparisons between the two-layer and the three-layer networks.

10.2. Recall of previously stored events: setting the threshold

Marr points out that the performance of the network is highly dependent on the method of setting the threshold during recall. He shows that if information is available about only the number of activated synapses or the proportion of activated synapses that were modified (but not both) then performance is much degraded. Information about both quantities is required. Our preliminary simulations confirmed this observation. The number of activated and modified synapses must exceed an absolute threshold, T, and the proportion of activated synapses that were modified must exceed a certain value, f. He suggests that a good way of manipulating the values of T and f during recall is to start with a low value of f and gradually increase it, whilst decreasing the value of T from a high value so as to maintain constant the total number of cells that are active. He left it to the reader to obtain a workable strategy. We have experimented with several different schemes, whose suitability for implementation depends on the type and amount of information that is available about the state of the net. Any biologically acceptable strategy should work for a wide range of values of parameters, such as the number of events, n.

1. Maximal similarity strategy

At each stage of recall the values of T and f were taken that cause the smallest number of cells in the given layer to be in the wrong state. This strategy may not be one that the nervous system could implement, but is probably close to the optimal strategy.

2. Staircase strategy

At a fixed, initially low, value of f, T is gradually decreased from a high value until the required level of activity is attained. f is then increased slightly, and T lowered further to restore the activity level. The process is repeated until either any further increase in f would prevent the activity from reaching the required level or f reaches the value 1.0.

3. Simple competitive strategy

The value of f is set at 0, and T lowered from a high value until the required level of activity is reached. This approximates the 'k-winner-take-all' competitive learning strategy, where k cells are required to be active.

To test recall capabilities, two tests were usually done. (i) Recall from subevents: the cues used were each a part (subevent) of a stored pattern. (ii) Recall from noisy cues: these cues were subevents of previously stored events to which a number of spuriously firing cells were added to make the total number of cells firing as in a full event.
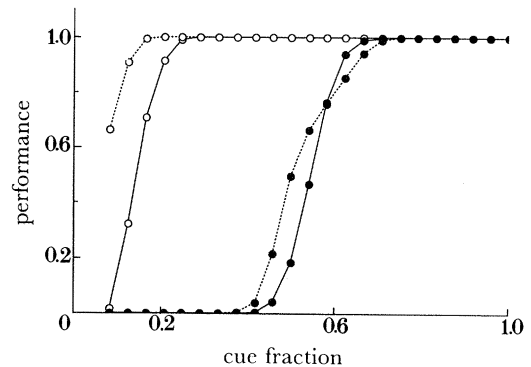
Figures 4–7. Performance of two layer model against the three layer model. In the simulation results shown in figures 4–7, unless otherwise stated, partial versions of the previously stored patterns were used as cues during recall to test the performance of these models. The number of active cells the cue shares with a stored pattern divided by the number of active cells in the stored pattern (cue fraction) is plotted on the horizontal axis. The vertical axis is a performance measure, the number of patterns recalled perfectly divided by the total number of patterns presented; (----), two-layer model and (———), three-layer model; (o), partial cue results; (●), noisy cues. The parameters were constructed such that the networks should store 1000 patterns with good recall performance, and in most cases 1000 patterns were stored. Figure 4 shows the performance of the two-layer model compared with that of the three-layer model. Canonical parameters, derived from Marr's constraints, are used. Retrieval from partial cues and from noisy cues is shown. The performance of the two models is comparable even though the three-layer model has many more synapses.
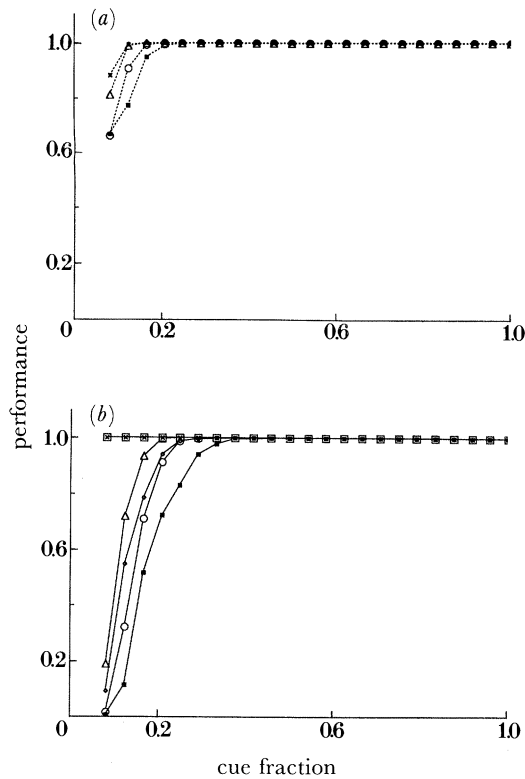
Figure 5. Effect on performance of using different numbers of connections. (*a*) Two-layer model. (○), canonical connectivity; (■), 20% fewer connections between layers; (△), 20% more connections; (×), full connectivity. (*b*) Three-layer model. (○), canonical connectivity; (■), 20% fewer connections between layers; (△), 20% more connections; (◇), canonical connectivity between $\mathscr{P}_1$ and $\mathscr{P}_2$ and full connectivity between $\mathscr{P}_2$ and $\mathscr{P}_3$; (□), full connectivity between $\mathscr{P}_1$ and $\mathscr{P}_2$ and canonical connectivity between $\mathscr{P}_2$ and $\mathscr{P}_3$; (×), both layers fully connected. More connections improve performance, especially when connections are added to the relatively sparse $\mathscr{P}_1$ to $\mathscr{P}_2$ projection.

11. SIMULATION RESULTS

Unless otherwise stated, in all simulations we used the first threshold setting strategy (maximal similarity). Figure 4 compares the fidelity of recall for the two nets as specified in §10. The performance of the two nets is comparable. To obtain perfect recall from noisy cues, the number of genuine cells must be significantly greater than for partial cues.

To check how far our results depended on a serendipitous choice of parameter values, we investigated the effects of varying the connectivity from the canonical values used in figure 4. Figure 5 shows the effects of varying the values of parameters S_2 and S_3, the number of connections made on a cell in $\mathscr{P}_2$ and $\mathscr{P}_3$, respectively. In general, the more connections there are, the better the performance. In the three-layer model, best performance is attained by having the $\mathscr{P}_1 \to \mathscr{P}_2$ layer fully connected; having $\mathscr{P}_2 \to \mathscr{P}_3$ fully connected is not so crucial. The performance for a three-layer net is slightly better than for a two-layer net. We also checked the effect of the collateral system on performance, and this does bring a slight improvement.
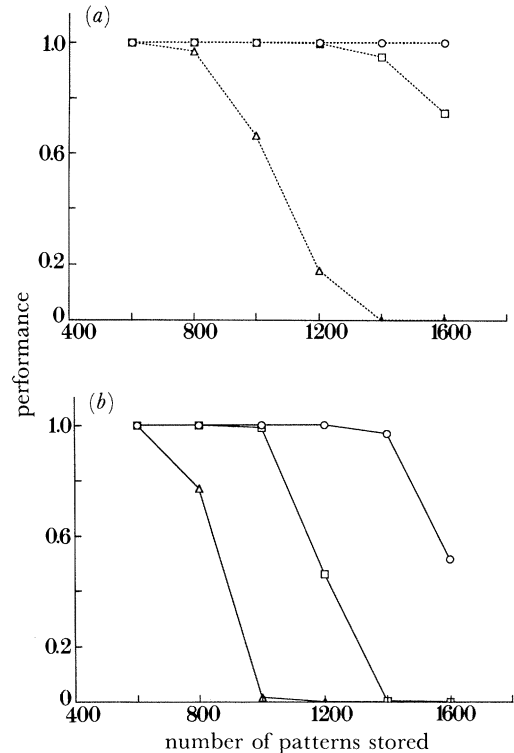

number of patterns stored

Figure 6. Effects of network loading. The performance for networks with 600, 800, ..., 1600 patterns stored. The number of patterns stored is plotted on the horizontal axis. Canonical parameters were used. Fraction of genuine bits in cue: (△), 0.08; (□), 0.25; (○), 0.5. (*a*) Two-layer model. (*b*) Three-layer model.

We also illustrate that the networks indeed function as content-addressable memories with a capacity of roughly 1000 patterns. Figure 6*a* shows that for the two-layer model, partial cues of 8% of the size of the full event result in 66% of the patterns being recalled perfectly. Partial cues of 25% full size gave perfect recall. (Figure 4 shows that 99% perfect recall is obtained with cues of 16% of the full size.) If more events are stored then larger cues are required. For the three-layer model, a cue of 25% full size is required to yield good performance (figure 6*b*).

The other facet of these two networks that we investigated was the effect of different threshold setting strategies. Figure 7*a* shows that for the two-layer net, the maximal similarity strategy yields a better performance than the staircase strategy, which itself is better than simple competitive learning. Corresponding results for the three-layer net are shown in figure 7*b*. The amount of knowledge required to be specified about the network is most for the strategy of maximal similarity and least for competitive learning.

12. CONCLUSIONS

Marr's model of the hippocampus as a temporary memory device bears a resemblance to mammalian hippocampus, in that both are multilayer, have some topographic ordering and a feedback loop. At a deeper level, the resemblance is less compelling, as the model lacks any representation of the details of the trisynaptic circuit, which were known at the time, and also details
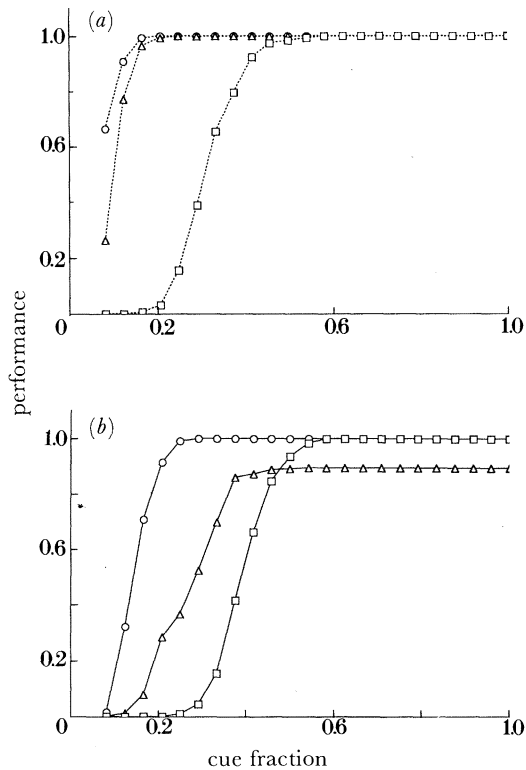
[115]

Figure 7. Effects of various threshold setting techniques. (○), maximal similarity; (△), staircase strategy; (□), competitve strategy. (*a*) Two-layer model. The maximal similarity and staircase strategies are comparable, both performing much better than the simple *k*-winner-take-all competitive strategy. (*b*) Three-layer model. Performance using the maximal-similarity strategy is better than that of the other two. Performance using the competitive strategy is comparable with its performance for the two layer model. The staircase strategy does not do as well as it did for the two-layer case. Analysis of the simulation data show that it does not elicit very good representations in $\mathscr{P}_2$ because of the relatively sparse $\mathscr{P}_1$ to $\mathscr{P}_2$ projection.

of other pathways, such as the direct projections from entorhinal cortex to CA1 and CA3, which were not. In our opinion, his computational reasons for deciding upon a three-layer model are invalid. Our simulations, albeit on a cut-down version of his model, suggest that the two-layer and three-layer models perform equally well.

Our evaluation of Marr's model has suggested two possible new lines of research.

Given the wealth of new anatomical information now available about the hippocampal formation, it would be possible to build a model of temporary memory, employing constraints similar to those used by Marr. However, answers to apparently elementary questions, such as the required capacity of the memory, are required to constrain the problem.

On the more cellular level, Marr made it clear that in incompletely connected networks, which are biologically more realistic than fully connected nets, dual threshold-setting strategies are essential. Such strategies could be investigated in terms of their capabilities, computational costs, and their physiological and neuroanatomical plausibility. For example, it is un-

likely that each memory cell could have knowledge available about the state of its own synapses, but rather that there is information available about the synaptic states of groups of cells.

REFERENCES

Andersen, P., Bliss, T. V. P. & Skrede, K. 1971 Lamellar organization of hippocampal excitatory pathways. *Expl Brain Res.* **13**, 222–238.

Blackstad, T. W. 1956 Commissural connections of the hippocampal region in the rat, with special reference to their mode of termination. *J. comp. Neurol.* **105**, 417–538.

Bliss, T. V. P. & Lømo, T. 1973 Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetised rabbit following stimulation of the perforant path. *J. Physiol. Lond.* **232**, 331–356.

Brindley, G. S. 1969 Nerve net models of plausible size that perform many simple learning tasks. *Proc. R. Soc. Lond.* B **174**, 173–191.

Cajal, S. R. 1911 *Histologie du système nerveux.* Tome II. Madrid: CSIC.

Cragg, B. G. 1967 The density of synapses and neurons in the motor and visual areas of the cerebral cortex. *J. Anat.* **101**, 639–654.

Churchland, P. S. & Sejnowski, T. J. 1989 Neural representation and neural computation. In *Neural connections, mental computation* (ed. L. Nadel, L. A. Cooper, P. Culicover & R. M. Harnish), pp. 15–48. MIT Press, Massachusetts: Bradford Books.

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.

Grossberg, S. 1973 Contour enhancement, short term memory, and constancies in reverberating neural networks. *Stud. app. Math.* **52**, 217–257.

Grossberg, S. 1987 Competitive learning from interaction to adaptive resonance. *Cog. Sci.* **11**, 23–63.

Hebb, D. O. 1949 *The organisation of behaviour.* New York: John Wiley.

Kohonen, T. 1972 Correlation matrix memories. *IEEE Trans. Comp.* **C-21**, 353–359.

Lømo, T. 1971 Potentiation of monosynaptic EPSP's in the perforant path – dentate granule cell synapses. *Expl Brain Res.* **12**, 46–63.

Lorente de No, R. 1933 Studies on the structure of the cerebral cortex. I. The Area Entorhinalis. *J. Psychol. Neurol. (Lpz).* **45**, 381–438.

Lorente de No, R. 1934 Studies on the structure of the cerebral cortex. II. Continuation of the study of the Ammonic system. *J. Psychol. Neurol. (Lpz).* **46**, 113–177.

Marr, D. 1969 A theory of cerebellar cortex. *J. Physiol., Lond.* **202**, 437–470.

Marr, D. 1970 A theory for cerebral cortex. *Phil. Trans. R. Soc. Lond.* B **176**, 161–234.

Marr, D. 1971 Simple memory: a theory for archicortex. *Phil. Trans. R. Soc. Lond.* B **262**, 23–81.

Marr, D. 1974 The computation of lightness by the primate retina. *Vision Res.* **14**, 1377–1388.

Marr, D. 1982 *Vision.* New York: Freeman.

Raisman, G., Cowan, W. M. & Powell, T. P. S. 1965 The extrinsic afferent, commissural and association fibres of the hippocampus. *Brain* **88**, 963–996.

[116]

Sejnowski, T. J., Koch, C. & Churchland, P. S. 1988 Computational neuroscience. *Science, Wash.* **241**, 1299–1306.

Squire, L. R., Shimamura, A. P. & Amaral, D. G. 1989 Memory and the hippocampus. In *Neural models of plasticity* (ed. J. Byrne & W. Berry), pp. 208–239. New York; Academic Press.

White, L. E. Jr 1959 Ipsilateral afferents to the hippocampal formation in the albino rat. I. Cingulum projections. *J. comp. Neurol.* **113**, 1–41.

Willshaw, D. J., Buneman, O. P. & Longuet-Higgins, H. C. 1969 Non-holographic associative memory. *Nature, Lond.* **222**, 960–962.

[117]