From the Retina to the Neocortex

Selected Papers of David Marr

Edited by Lucia Vaina

With Commentaries by

Jack D. Cowan
W. Eric L. Grimson
Norberto M. Grzywacz
Ellen C. Hildreth
Bruce McNaughton
Terrence J. Sejnowski
W. Thomas Thach
David Willshaw

Birkhäuser
Boston · Basel · Berlin 1991

SIMPLE MEMORY: A THEORY FOR ARCHICORTEX

By D. MARR
Trinity College, Cambridge

(*Communicated by G. S. Brindley, F.R.S.—Received* 27 *July* 1970—*Revised* 12 *November* 1970)

CONTENTS

Vol. 262. B. 841. 3 [Published 1 July 1971

It is proposed that the most important characteristic of archicortex is its ability to perform a simple kind of memorizing task. It is shown that rather general numerical constraints roughly determine the dimensions of memorizing models for the mammalian brain, and from these is derived a general model for archicortex.

The addition of further constraints leads to the notion of a simple representation, which is a way of translating a great deal of information into the firing of about 200 out of a population of 10^5 cells. It is shown that if about 10^8 simple representations are stored in such a population of cells, very little information about a single learnt event is necessary to provoke its recall. A detailed numerical examination is made of a particular example of this kind of memory, and various general conclusions are drawn from the analysis.

The insight gained from these models is used to derive theories for various archicortical areas. A functional interpretation is given of the cells and synapses of the area entorhinalis, the presubiculum, the prosubiculum, the cornu ammonis and the fascia dentata. Many predictions are made, a substantial number of which must be true if the theory is correct. A general functional classification of typical archicortical cells is proposed.

0. Introduction

The cortex of the mammalian cerebrum admits a crude division into two classes: the archicortex, which is relatively simple and primitive; and the neocortex, which has developed more recently and is very elaborate, especially in man. In a recent paper (Marr 1970), a general theory for neocortex was set out. The present paper provides its counterpart for archicortex.

The comparatively simple structure of archicortex is probably reflected in its performance of a comparatively simple function. The central point of the neocortical theory was that a particular method of organizing information is likely to be useful in many different circumstances: it was shown how neocortex might take advantage of this to change the language in which incoming information is expressed by reclassifying it, as well as carrying out routine storage of associations between existing classes. It will be argued in the present paper that archicortex cannot reclassify information in this way. It will be shown that its histology is consistent with the proposition that it performs only a simple memorizing function—storing information in the language in which it is presented—rather than with organizing information in any more complicated sense. Recent work on the storage of information in nerve nets (Brindley 1969; Marr 1969, 1970) has reduced the construction of such a theory to little more than a technical exercise: it is an unavoidable one none the less, and various interesting factors emerge from this study.

The paper consists of three main divisions. In the first, §§ 1 and 2, the main ideas behind simple memory theory are explained. These ideas lead to a particular neural model which, it is proposed, captures the essence of much of the archipallial cortex. It is shown that under certain circumstances, the performance of such a model can be greatly improved by use of collateral synapses between its cells (the *collateral effect*, § 2.4).

The second part of the paper, § 3, takes an explicit model constructed along the lines suggested by the first part, and derives the equations which describe its expected performance. The model's storage capacity and recall abilities for a selection of values of the important parameters are displayed in a number of tables. The computations (§ 3.1) are followed in the rest of § 3 by a rough justification of the values of the parameters chosen.

The third part of the paper (§ 4) uses the model of § 3.1 to arrive at a theory of the hippocampal cortex. This theory produces many testable predictions, which are summarized in § 5. The theory is restricted to operations within the cortex, and does not describe any input–output relations. The reason is that they are much more complex than, for example, those of the cerebellar cortex, and their inclusion in this paper would have made it prohibitively long. They will therefore be set out elsewhere, together with the necessary extra theory.

0.1. *Notation*

Many of the terms and symbols of Marr (1970) are used in this paper, and it is convenient to repeat their definitions here. A *fibre* (e.g. $a_i(t)$) is a function of discrete time $t\,(=0, 1, 2, \ldots)$ and has the value 0 or 1. An *event* on the set $A = \{a_1, \ldots, a_N\}$ of fibres assigns to each fibre a value 0 or 1. Letters like E, F are used for events, and the value that E assigns to the fibre a_i is written $E(a_i)$. The phrase 'a_i in E' means 'a_i takes the value 1 in the event E'. A *subevent* on the set $A = \{a_1, \ldots, a_N\}$ of fibres is an event on a subset of A. Letters like X, Y denote subevents; and the set of fibres to which X assigns a value is called the *support* of X, and is written $S(X)$. Gothic letters like $\mathfrak{E}, \mathfrak{F}$, denote collections of events; and letters like $\mathfrak{X}, \mathfrak{Y}$ denote collections of subevents.

The event E is said to be a *completion* of the subevent X, written $E \vdash X$, if E and X agree at all the fibres to which X assigns a value.

Let $\mathfrak{E}$ be the space of all events over $\{a_1, \ldots, a_N\}$. An *r-codon* c on $\mathfrak{E}$ is a function, taking the values 0 or 1, such that $c(E) = 1$ if and only if a particular subset of r fibres $(a_{i_1}, \ldots, a_{i_r})$ all have the value 1 in E; c may be regarded as a detector of the subset $(a_{i_1}, \ldots, a_{i_r})$. An (R, θ)-*codon* is a similar function c such that $c(E) = 1$ if and only if at least θ of a particular collection $(a_{i_1}, \ldots, a_{i_r})$ of fibres have the value 1 in E.

1. General constraints

1.0. *Introduction*

It has recently been argued that neocortex may be regarded as a structure which classifies the information presented to it (Marr 1970). The detectors of the classes it forms are the pyramidal cells of layers V, III and possibly also of layer II. An incoming signal will probably pass through many such classifications during the course of its analysis. The number through which it passes will depend upon the animal, and upon its interest in that kind of information at that moment: it is clear that information is often abandoned as uninteresting before it has been examined to the maximum depth of which the animal is capable.

It is probably reasonable to suppose that at a given moment, there will exist in an animal's brain information whose expression is now as sophisticated as the animal either requires, or can provide. Further classification of the information may be carried out later but, at that moment, the animal needs simply to be able to store it in its present form. Such an expression of the input is called the animal's *current internal description* of the environment, and it is the storage of the current internal description which constitutes the animal's memory of the information. From these memories, he will form new classificatory units, organize temporally extended actions, and arrange to respond in the appropriate way to pieces of subsequent current internal descriptions.

The problems that are studied in this paper are those which arise in the storage and the free association of such current internal descriptions. The central problem may, by the neocortical theory (Marr 1970), be translated into the following form. $\mathscr{P}$ is a large population of neocortical pyramidal cells, of which some are firing. It is required that this should be recorded in some way, so that firing in a few of the cells which are active together in some event E can later elicit the firing of all cells active in E. This scheme is probably only remotely analogous to hippocampal input-output relations in most mammalian brains, but it is a convenient model with which to introduce the cortical theory.

Three considerations necessitate the construction of a special theory for this problem. First, although it has been shown that the neocortex can store associations between classificatory units

(Marr 1970, §4)—for example through the pyramidal cells' basilar dendrites—this kind of storage requires a rather special kind of pre-existing structure: the relevant fibres have already to be distributed to roughly the correct places. Direct storage of associations in this way makes heavy demands on the abundance of interconnexions.

The second consideration concerns the way this kind of associational storage works. It essentially involves recording at each active pyramidal cell Ω_i in $\mathscr{P}$ a list of many of the cells Ω_j co-active with Ω_i. This can become very expensive, and there are ways of improving upon it. Furthermore, it is only worth recording information in a permanent memory when it is known fairly certainly how that information should be expressed. It may, for example, turn out that part of a current internal description should be recoded to form a new classificatory unit. If this were done, a direct associational storage of that current internal description would soon be obsolete: it is better to store it temporarily in a special associative memory, until it becomes quite clear how it should be permanently set down.

Thirdly, there are many instances in which the control of behaviour would be made rather easy if an associative memory were available as a temporary storage place for instructions. This facility would, for example, allow an instruction of the form 'see post-box–post letter' to be set up before one started out on a walk.

1.1. *Simple memory*

Let $\mathfrak{E}$ be the set of all events and all subevents on the fibres $\{e_1, e_2, \ldots, e_m\}$, and let $\mathfrak{F}$ be the set of all events on the fibres $\{f_1, f_2, \ldots, f_n\}$ (see §0.1 for definitions of these terms). As time t progresses ($t = 0, 1, 2, \ldots$), denote the event at time t in $\mathfrak{E}$ by E_t, and that at time t in $\mathfrak{F}$ by F_t. A *simple memory* is a device which connects E_t and F_t, for each t, in the following sense. Let X be a subevent or an event in $\mathfrak{E}$. Let $X_1, \ldots, X_J$ be all the completions of X in $\mathfrak{E}$; that is $X_i \vdash X$ for $1 \leqslant i \leqslant J$, and there are no others. (If X is an event, its completion is unique and is itself.) Suppose that exactly one of the events $X_1, X_2, \ldots, X_J$ has occurred. That is, the equation $X_j = E_t$ has exactly one solution, for all values of j, and of t up to the present time. Then $\mathfrak{E}$ and $\mathfrak{F}$ are joined by a *simple memory* if presentation of X subsequently causes the event F_t in $\mathfrak{F}$.

Two special cases deserve separate names. In the case where $\{e_1, \ldots, e_m\} = \{f_1, \ldots, f_n\}$, the memory described above is called a *free simple memory*: if the memory is not free, it is called a *directed simple memory*. The reason for these names is that in a free simple memory, there are no constraints upon the way the associations may flow. Any collection of fibres from the set $\{f_1, \ldots, f_n\}$ may be used to recall the activity of the rest of these fibres at a particular time. In directed simple memory, this is not so. For example, f_1 may not be included in $\{e_1, \ldots, e_m\}$, in which case information about f_1 can never be used to recover information about the rest of the $f_i (2 \leqslant i \leqslant n)$.

In the models that are studied in this paper, rather little is said about whether

$$\{e_1, \ldots, e_m\} = \{f_1, \ldots, f_n\}.$$

The question is unimportant until the problem of input–output relations is studied. It is enough to note here that the same basic memory mechanism can be used for both free and directed simple memories.

1.2. *Numerical constraints*

There are various arguments which roughly determine the shape of simple memory theory; they are best presented in the form of order-of-magnitude calculations. This section contains four such arguments: the first is concerned with the proportion of learned to possible input events; the second with the likely size of input vocabulary—i.e. the number of input fibres; the third with the number of events which have to be held in the memory; and the fourth with the proportion of cells of the population concerned with the storage that is used for each event.

1.2.1. *The constraint of a limited history*

The number of fibres that may be involved in a current internal description must be expected to be quite huge; but even if it were only 1000, and a mere 10 were involved at each unit of time (say 1 ms), there are enough possible events for the system to run for more than 10^{12} years without repetition. The world is, of course, not random; but the figures 10 and 1000 are certainly underestimates. From this observation follow two conclusions. First, information about the current internal description concerns *whether* a particular event has occurred, rather than *how often* it has done so, since the answer to the latter question is almost certainly never or once. Secondly, very few of the possible events will ever actually occur. Recovery of an event will therefore be theoretically possible from an extremely small amount of information, and the design of neural models must be such as to allow this.

1.2.2. *Cortical indicator cells*

It is supposed that neocortical pyramidal cells of layers III and V are output cells for classificatory units, and that some, though not necessarily all, of such cells can take part in a current internal description. The human cerebrum contains about 7×10^9 cells (Shariff 1953) of which at least say 10^8 could be classed as cortical pyramids. This is a huge number, and any attempt to allow all the cells in a population of this size to have access to a simple memory would lead to an unacceptably large neural structure for that memory. If, however, the memory is used for a relatively small number of events (of the order of 10^5), information then being removed to the neocortex, an important simplification can be made.

Suppose that scattered more or less uniformly over the cerebral neocortex were cells which responded simply to activity in their neighbourhood of the cortex. If such a cell were driven by a very small region of the cortex—an area of perhaps $0.03\,mm^2$—it would serve as a marker of activity in the cortical pyramids within that region. Each cortical pyramid represents a separate classificatory unit, and it can probably be assumed that within such a region not all the pyramids will be active simultaneously. The non-specific cell which marks activity in that region is called an *indicator cell*: the best design for such a cell would probably assign to it a thin, unbranched ascending dendritic stem which passes through all layers of the neocortex, and which is sensitive to excitatory influences throughout its length.

The great advantage of indicator cells is that they can be used as entry fibres to a simple memory, provided that the return fibres synapse with the true cortical pyramids and not with the indicators. In this way, whenever a pyramidal cell is used, its nearby indicator(s) cause an entry to be made to the memory, while the return synapses to the pyramid itself are modified. The memory can later use these synapses to drive the original pyramidal cell. The only disadvantage arises when two nearby pyramidal cells are used in two different but very similar situations, but this problem is not a severe one.

A density of 30 indicator cells/mm^2 allows a quite sensitive specification of location; and although this figure is only a guess, we shall see in §3.1 that it can be changed by a factor of 10 without much disruption of the models analysed there. In general, the density of such cells should reflect the frequency with which the various regions of neocortex use the simple memory facility, the density being high in regions expressing information which often needs temporary storage, and low elsewhere. If indicator cells are used, one would expect their dendritic design to vary as well, being very compact in areas where their cell density is high, and perhaps arborizing where they are rare.

The total area of one hemisphere of the human cerebral cortex is estimated to lie between 800 and 1300 cm^2. If it is supposed that about 400 cm^2 need to have access to the simple memory (this figure may be too large), the memory will possess about 10^6 afferent fibres. This is the approximate number of fibres needing free simple memory, and does not include the various kinds of directed simple memory which may, for example, be involved in the planning of temporally extended actions.

1.2.3. *Capacity requirements*

The design of a memory requires some idea of the number of events to be stored, and of the amount of information from which recovery of a whole event should be possible. These two factors are linked, since if a memory has to be capable of recovering events from a very small amount of information, its capacity is much smaller than if most of the original event can be used to initiate recall. It is necessary to make a rough estimate of both requirements.

Simple memory has many uses, and the brain probably employs different structures for each use, though the structures are likely to conform to the same basic plan. For directed simple memory, it is very difficult to provide even a rough guess at the storage requirements. For free simple memory (an explicit model for which is developed in §3.1), some idea of the necessary capacity can be obtained. The figure will not be very high, since it is part of the general theory that information is moved out of the simple memory when it is known how best to do this. The two possibilities for the re-storing of the information currently in simple memory are (i) that it is moved to neocortex in the form of new classificatory units (see Marr 1970, §§4, 5); (ii) that it is moved to neocortex in the form of associations between existing classificatory units (through, for example, the basilar dendrites of neocortical pyramidal cells).

It has been suggested that at least a part of the transfer between simple memory and the neocortex takes place during sleep (Marr 1970, §5). This implies that simple memory must have adequate capacity for holding the events of at least one day. There are 86 400 s in 24 h, and although many events will not be moved out for some time, one probably does not store a new event every second. The figure of 10^5 is therefore taken as the kind of capacity required of the free part of the simple memory.

The amount of information which can recall an event is even harder to estimate, but it should probably be very small, less than a tenth of the information contained in the original event. The model of §3.1 operates at a level considerably below this figure.

1.2.4. *The activity of a collection of cells*

Let $\mathscr{P}$ be a population of M cells, $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_M$. Suppose that at time t, exactly L of these cells are firing: then the *activity* of $\mathscr{P}$ at time t is defined to be L/M, and is written $\alpha = \alpha(t)$.

If $\mathscr{P}$ is being used to store n input events, and if its activity during each is α, then each cell of

$\mathscr{P}$ may expect to be used in αn input events. If the storage is taking place in the cells of $\mathscr{P}$, each cell will have to learn part of about αn input events. The number of subevents a single cell can learn is determined by the number of modifiable afferent synapses it has, and by the number that are used in each subevent. For example, the number of fairly dissimilar events that a cerebellar Purkinje cell can learn is probably about 200 (Marr 1969). Purkinje cells have more afferent synapses than any cortical cells, and so it follows that most cortical cells will not be able to learn substantially more than 200 subevents. The number of input events that the population $\mathscr{P}$ described above may learn is therefore bounded by about $200\alpha^{-1}$. This is an important and rather general constraint.

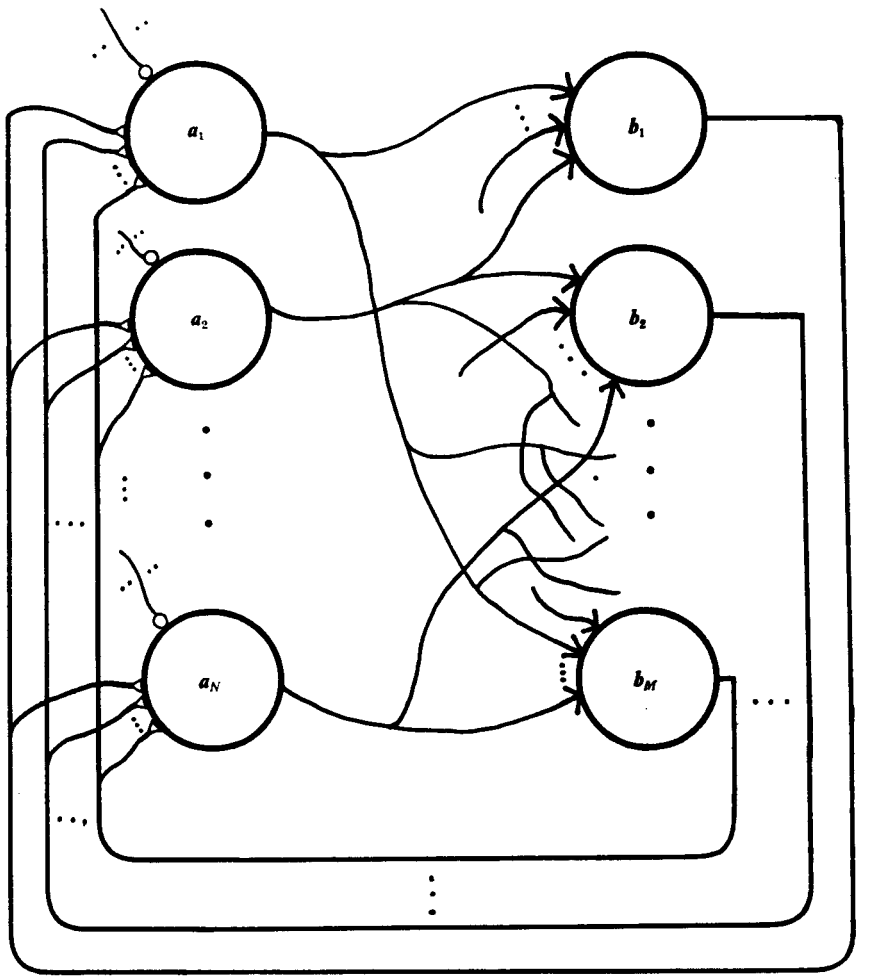

FIGURE 1. A primitive associative memory. The current internal description is an event on the cells $a_1, \ldots, a_N$: this is given a codon representation in the cells $b_1, \ldots, b_M$ (which have Brindley afferent synapses), and the return to the a_i-cells is through Hebb modifiable synapses. The various inhibitory interneurons necessary for the correct operation of the system have been omitted. This class of model provides an efficient associative memory for events on the a_i as long as their number and size are suitably restricted.

1.3. *The form of the analysis*

The model of figure 1 shows almost the simplest design for a free simple memory for events on the set of fibres $A = \{a_1, \ldots, a_N\}$. This model may be derived most quickly as follows. Let X be a subevent on A. Then the problem of recovering the completion E of X (assuming that exactly one such E has occurred) may be regarded as the problem of diagnosing those a_i with $E(a_i) = 1$ from

the information contained in the subevent X on the basis of the information stored in the memory. It is now possible to apply the interpretation theorem (Marr 1970, §§2, 4) to the problem, and figure 1 contains one arrangement for applying the corresponding neural analysis.

The inputs $a_1, \ldots, a_N$ are the cells which constitute the vocabulary of the current internal description, and the cells $b_1, \ldots, b_M$ are suitable evidence cells. The technique of codon formation is used to construct suitable evidence cells (see Marr 1970, § 4), and for this reason, the b_j afferents end in Brindley synapses. (*Hebb* synapses will be taken to mean synapses that are initially ineffective, but are facilitated by simultaneous pre- and post-synaptic activity. *Brindley* synapses are Hebb synapses that also contain an unmodifiable excitatory component (Marr 1970, § 4.3.1; Brindley 1969).) The b_j-cell population contains appropriate threshold-setting inhibitory interneurons, whose function is to keep the number of b-cells that are active roughly constant during both storage and recall. These interneurons do not appear in the figure.

The return projection to the a-cells ends in Hebb synapses. There are inhibitory interneurons in the a-cell population which, during recall, allow firing in only those a-cells the highest proportion of whose active afferent synapses from the b-cells have been modified. This corresponds to implementing the interpretation theorem at the a-cells, in response to the subevent X. The cell a_i measures $\mathcal{P}(a_i|X)$ when X is applied to the set $\{a_1, \ldots, a_N\}$ (Marr 1970, § 2.5), and the b-cell thresholds are lowered in such a way as to keep the number of b-cells that are firing roughly constant (Marr 1970, § 4.4).

In principle, free simple memory is obtained by allowing the projections from the a-cells to the b-cells and back to be distributed freely over both populations (as in figure 1). A directed simple memory is obtained, for example, by arranging that only certain a-cells project to the b-cells, and that only certain a-cells receive projections from the b-cells.

1.4. *The consequences of the numerical constraints*

In this section are outlined the principal effects of the constraints described in § 1.2 when they are applied to the kind of model to which the methods of § 1.3 give rise. The development is informal, and is designed to give the reader an overall view of the theory developed in §§ 2 to 4. Its main purpose is to show roughly why it is that the basic model of figure 1 is inadequate for simple memory, and how this leads to the idea that a special working representation of each input has in fact to be created in the memory. This central representation is a kind of template for each event; it probably involves rather few cells—perhaps only 100 to 1000 even in man—and provides an economical central storage pattern from which the event in the output space $\mathfrak{F}$ at that particular instant can be recovered. This representation, called the *simple representation* of the current internal description, is a central feature of the present paper.

1.4.1. *Synaptic modification*

Where codon formation occurs, the relevant synaptic modification has been regarded as an all-or-none process (Marr 1970, § 4). In contrast, the afferent synapses to output (cortical pyramidal) cells need to have variable strength in order to measure $P(\Omega|c_i)$, although it may be that this is in practice approximated by an all-or-none process (Marr 1970, §§ 4, 7). The numerical constraints of § 1.2 imply that in the theory of archicortex, synaptic modification should probably be regarded as an all-or-none process, although it is allowed that different classes of synapses may have different maximum strengths.

One reason for this is as follows. For evidence cells (i.e. in codon formation) the arguments are the same as for neocortex: these synapses are involved in representing a diagnostic space, not in measuring probabilities therein. For diagnostic processes in a simple memory, the argument rests on the peculiar way in which the memory is used—as a temporary store to which new information is continually being added. At a neocortical output cell, the notion of a conditional probability has a practical meaning, since the output cell and its supporting evidence cells are structures which form a permanent part of the brain's interpretive apparatus. This is not true of simple memory. Much of the information held therein is needed only temporarily, and that which is not will be removed to the neocortical store when it becomes clear how it should be represented there. The notion of conditional probability in such circumstances has at best only a changing meaning.

1.4.2. *Inadequacy of the simple model*

It is easy to show by using order-of-magnitude calculations that the simple model of figure 1 cannot be applied to the case where there are as many as 10^6 input cells $\boldsymbol{a}_i$. Since neocortical pyramidal cells probably possess fewer than 100000 afferent synapses, most of which will be occupied with standard diagnostic evidence and with permanent neocortical associative information, it can probably be assumed that only about 10^4 synapses are available for the simple memory function. In the simple model outlined in figure 1, this means that the number of $\boldsymbol{b}$-cells, M, may be taken as 10^4, each one synapsing with every one of the 10^6 $\boldsymbol{a}$-cells. The $\boldsymbol{b}$-cells must possess modifiable synapses since, otherwise, recall from subevents of learnt events would be impossibly bad. If the capacity of the memory is taken to be about 10^5 events, and each $\boldsymbol{b}$-cell can learn 10^2 (§ 1.2), the activity α of the $\boldsymbol{b}$-cell population must be as low as 10^{-3}—that is, 10 cells active at any instant. This number is too small to allow a reliable representation of the whole input event by the $\boldsymbol{b}$-cells, and the model is therefore inadequate.

1.4.3. *The simple representation of the current internal description*

Arguments like that outlined in § 1.4.2 show two things: first, that there must be more than one layer of cells (like the $\boldsymbol{b}$-cells) between the input and the return of a simple memory, if it is bound by numerical constraints like those described in § 1.2. Secondly, the small number of synapses available at neocortical pyramids for the simple memory means in effect that there will be rather little spare capacity in the projection back from the simple memory. That is, most of the storage capacity at these synapses will be exhausted by the straightforward task of relating the pyramids to the activity in the projection from the memory during full events: there will be little left over to help in the task of completing a subevent of a learnt event. This means that during recall of a learnt event from a subevent, the recall must have been virtually achieved by the memory *before* the signals reach the projection back to the neocortex. Hence most of the diagnostic analysis involved in discovering the completion of a subevent takes place in the memory itself, not at the $\boldsymbol{a}$-cells. In the simple case of figure 1 (which can be used to store rather few events), this would mean that a subevent X of E could recall E only if it caused activity in the same $\boldsymbol{b}$-cells as did E.

This is a rather stringent condition on the structure of the memory. It means that there exists a stage—a layer of cells—in (and by) which the completion process is achieved. Each input event E has a representation as a firing pattern in this population of cells, and the problem of completing

a subevent X of E is equivalent to the problem of recovering its corresponding firing pattern. This pattern is called the *simple representation* of the input E.

1.4.4. *Advantages of the simple representation*

The notion of the simple representation of an event E of the current internal description makes many of the problems of free and directed simple memory easy to express. A simple representation needs to be formed only of those parts of E that contain the subevents through which E will later be addressed: and the simple representation needs to be associated back (through the return from the memory) only to those parts of E that will need to be recalled.

It will turn out that simple representations consist of collections of cells in a population whose activity α(§ 1.2.4) is very low. The activity is in fact so low ($\alpha \approx 0.001$) that the cells of a simple representation can be directly associated to each other by collaterals terminating in Hebb synapses. The simple representation of E, written $[E]$, can thus be regarded as a firing pattern which can complete itself through its collateral synapses (called the *collateral effect*, § 2.4). Again, simple representations are somewhat limited in the maximum size they can attain, and this leads to the notion that more than one simple representation may be formed, each dealing with a different subevent of E. Within each simple representation, there is a full collateral effect, but between any two, it is less full (see § 4.5.1).

2. The basic model for archicortex

2.0. *Introduction*

The arguments of § 1 show that simple memory may be divided into two operations: the creation of suitable diagnostic spaces for the input events as they occur; and the performance, during recall, of diagnostic operations within those spaces. The representation of these two basic functions requires a model consisting of two parts, closely analogous to codon formation and output cell selection in the neocortical theory. Many of the factors which determine the shape of each component have already arisen in the theory of the neocortex: they can therefore be derived rather quickly, and with this the first two parts of this section are concerned.

Within the outlines established by these two basic models, the actual shape of a simple memory is determined largely by numerical constraints. The rest of this section therefore shows how the capacities and characteristics of various models may be calculated, and derives the conditions imposed by the fact that the cells involved have to be physiologically plausible.

2.1. *Codon formation*

The first task to be discussed is the construction of evidence functions by input events. The obvious way to do this is to use the technique of codon formation, described in some detail by Marr (1970, § 4.3). (Compare also the *s*-cells of Brindley 1969.) The basic models for this appear in figure 2, and the arguments for each will be set out here only in so far as they differ from those put forward in the neocortical theory.

2.1.1. *Preference for the model 2 using Brindley synapses*

The main differences between the arguments appropriate here and those for the neocortex arise because the function of simple memory is to record *all* its incoming information: the difficulties which arose in the neocortex, concerning the formation of evidence only over the appropriate

diagnostic space, do not arise here. Model 1 of figure 2 is excluded for the same reasons as in the neocortical theory: each cell can represent only one event, since after one modification, all the synapses not used in that event become ineffective. Model 3 is excluded for two reasons: (*a*) a climbing fibre system cannot both be simple and choose those cells most appropriate for each event (i.e. those at which the greatest number of active afferents have synapses); (*b*) a climbing fibre system in any case requires more cells than model (2).

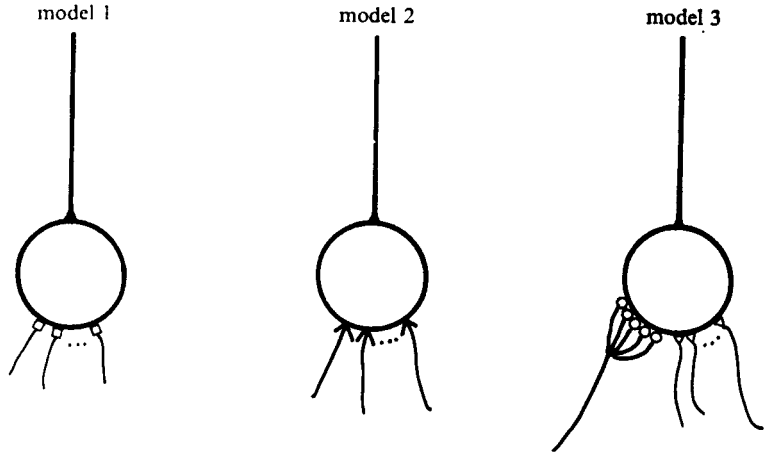

FIGURE 2. Three models for codon formation: model 1 uses synapses which are initially excitatory, but become ineffective as a result of post- without pre-synaptic activity; model 2 uses Brindley synapses; model 3 uses a climbing fibre and Hebb synapses.

2.1.2. *Threshold setting in model* 2

Brindley synapses contain an unmodifiable excitatory component, and are facilitated by a combination of pre- and post-synaptic depolarization. The post-synaptic threshold for the existence of modification conditions there will have to vary for two reasons: first the number of active afferents will not be constant; and secondly the overall proportion of synapses that have been modified will change, thus changing the amount of post-synaptic depolarization that an unlearned input of fixed size may be expected to cause. These problems do not arise in the special case considered by Brindley (1969), where the number of active afferents is always two, and the ratio of modifiable to unmodifiable components in the synapses is 1:2.

Synaptic modification probably depends on the local conditions prevailing in a piece of dendrite, and hence inhibition intended to prevent these conditions from arising must be applied directly to the dendrite. The use of Brindley synapses in codon formation therefore requires that inhibition of the appropriate strength should be applied to the dendrites containing those synapses.

There are broadly speaking two methods of providing such inhibition: either it is done by inhibitory cells which are otherwise identical to the codon cells; they learn inputs at the same rate, and are therefore excited at a rate which increases with the number of learnt events: or a negative feedback system is used, built to keep the number of codon cells that are active roughly constant. The first scheme is probably unsatisfactory, and the second is embodied in the model of figure 3. This model contains two kinds of inhibitory influence on the codon cell dendrites (often through different dendrites of the same inhibitory cell—e.g. the *G*-cells). One influence,

the inhibition driven directly by the afferent fibres, sets the cell thresholds on the assumption that no synapses have been modified. The other, a negative feed-back driven by codon cell axon collaterals via the *G*-cells, provides the component required to counteract the extra excitation which arises because a fraction of the population of synapses will have been modified by previous events. The system is imagined to be constructed so as to maintain a constant activity α in the set $\mathscr{P}$ of codon cells. The effect of all inhibition described here is subtractive, and dendritic branches which are not close are imagined to be independent.

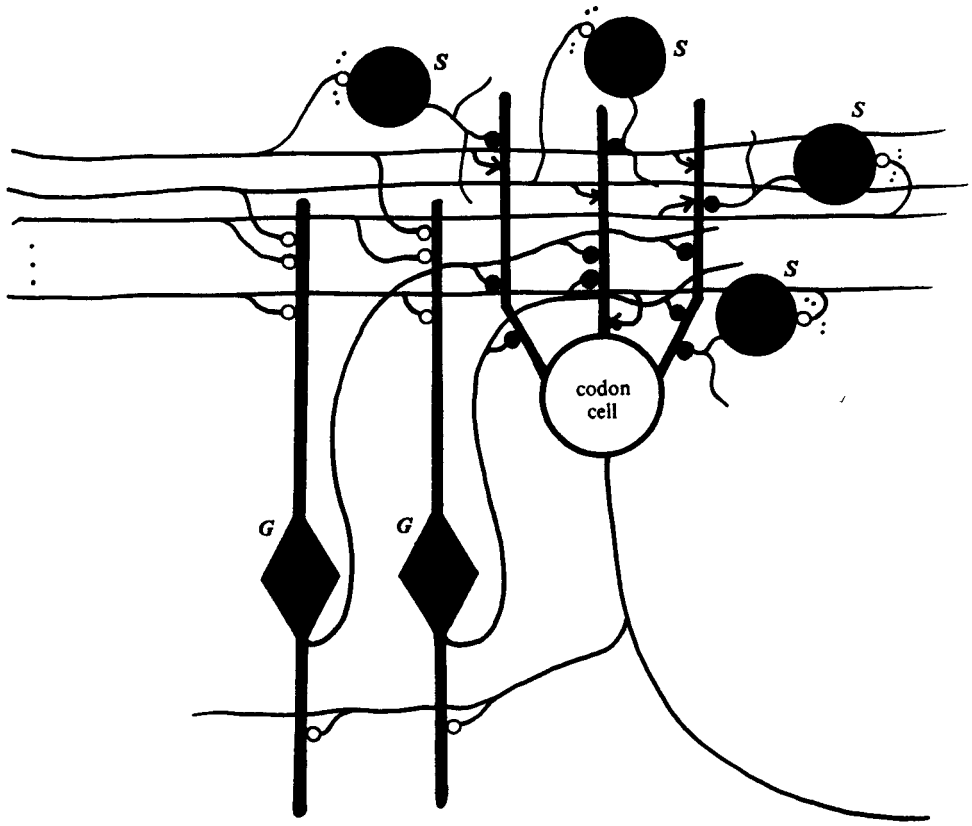

FIGURE 3. The full model for codon formation using Brindley synapses. Modification conditions are decided locally in the codon cell dendrites, and hence inhibition which controls these conditions is itself applied to the dendrites. The *S*-cells, driven by codon cell afferents, subtract roughly the expected excitation due to the unmodifiable component of the Brindley synapses. The *G*-cells, driven in part by codon cell axon collaterals, use negative feedback to compensate for changes in the size of the input event, and in the number of synapses which will already have been modified.

G. S. Brindley (personal communication) has pointed out that the need for *G*-cells in codon formation evaporates if information decays in the memory at about the same rate as it is acquired.

2.1.3. *Recalling an event*

The recall of an event is initiated by addressing the memory with a subevent. In order to avoid the problem of how the memory knows whether to store a given input, or to use it to recall the event most like it, it will be assumed that events which are to be stored are much larger than the

subevents which initiate recall. The reason for making this assumption is that the effect of a small subevent on the dendrites of the codon cells may then be regarded as being too mild to provoke synaptic modification there, since synaptic modification presumably requires a rather severe dendritic depolarization. The more general problem of controlling when a memory does and does not store its inputs will be dealt with in the paper on input–output relations.

One other point is needed to complete the discussion of codon cells. If the subevent that is being used for recall is wholly contained in the event to be recalled, then the best strategy is to lower the codon cell threshold until about the usual number of cells becomes active. This step is part of the usual procedure for implementing the interpretation theorem (Marr 1970, §§ 2.5 and 4.4). If, however, the subevent is only partially contained in the event to be recalled, then it will be shown in § 3.1 that better results are obtained if codon cells are treated like output cells (see § 2.2). This is essentially because output cells (with afferent basket synapses) are regarded as being capable of performing a division (Marr 1970, § 4.1.6); and, in the second situation, it turns out that the fraction of active afferent synapses which have been modified is a more suitable measure than the absolute number of such synapses.

2.2. *Diagnosis in simple memory*

It has been argued informally (§ 1.4.3) that the recall process in a simple memory has to be virtually complete by the time information is returned to the neocortical pyramidal cells. This means that the memory must contain internal diagnostic structure capable of recovering the pattern of firing appropriate to the learnt event of which the current input subevent formed a part. In this section, the cells at which the recovery is performed are described.

2.2.1. *The simple representation*

In the neocortical theory, it was imagined that information was represented by a family of classes, each of which was formed because of a clustering of input subevents. The function of simple memory is to record information as it occurs, without trying to produce the best possible classification of the input on the spot. It is proposed that information in a simple memory is also represented by a family of classes, but that in this case, the classes are chosen randomly. An incoming event is assigned to a family of cells, analogous to neocortical output cells, chosen because they happen to have more relevant synapses than any others. These cells may be regarded as 'random' variables taking the value 0 or 1: the probability that they have the value 1 is assessed at each moment by consulting the relevant evidence, in the usual way.

When viewed as random classes in this way, it is seen that the diagnosis and interpretation theorems may be applied to the assessment of the incoming evidence: indeed, these results, strictly speaking, are more accurately applied to the problem of the diagnosis of random classes than of the more organized objects for which they were developed (Marr 1970, § 2). Since it is assumed that modifiable synapses for simple memory have all-or-none modification characteristics, it follows that they should transmit a measure of the fraction f of their active afferent synapses which have been modified, provided that f exceeds some (variable) lower bound p (say).

It is thus proposed that the simple memory sets up, by a more or less random process, a set of classes which is unique (with very high probability) to each input. Each class is represented by a separate cell, although a given cell may represent more than one class. The set of cells which represent a given input in this way is called the *simple representation* of that input. The recall of an

event from a subevent is performed by recovering those classes by which the subevent is best interpreted, in the sense of the interpretation theorem (Marr 1970, § 2.5). In order to do this, the cells involved in a simple representation need to be able to measure the fraction f defined above.

2.2.2 *Output cells for a simple representation*

The theory of output cells for the random classes described in § 2.2.1 falls into two parts: the first describes the formation of the classes, and the second deals with the subsequent interpretation of inputs. The idea that these cells do two things—i.e. store and interpret—and that they do both things all the time, leads naturally to the question of how they know what to do to a given input. For now it is enough to assume that if an input is a subevent of a previously learnt event, it will automatically cause recall of that event. If not, it is simply stored.

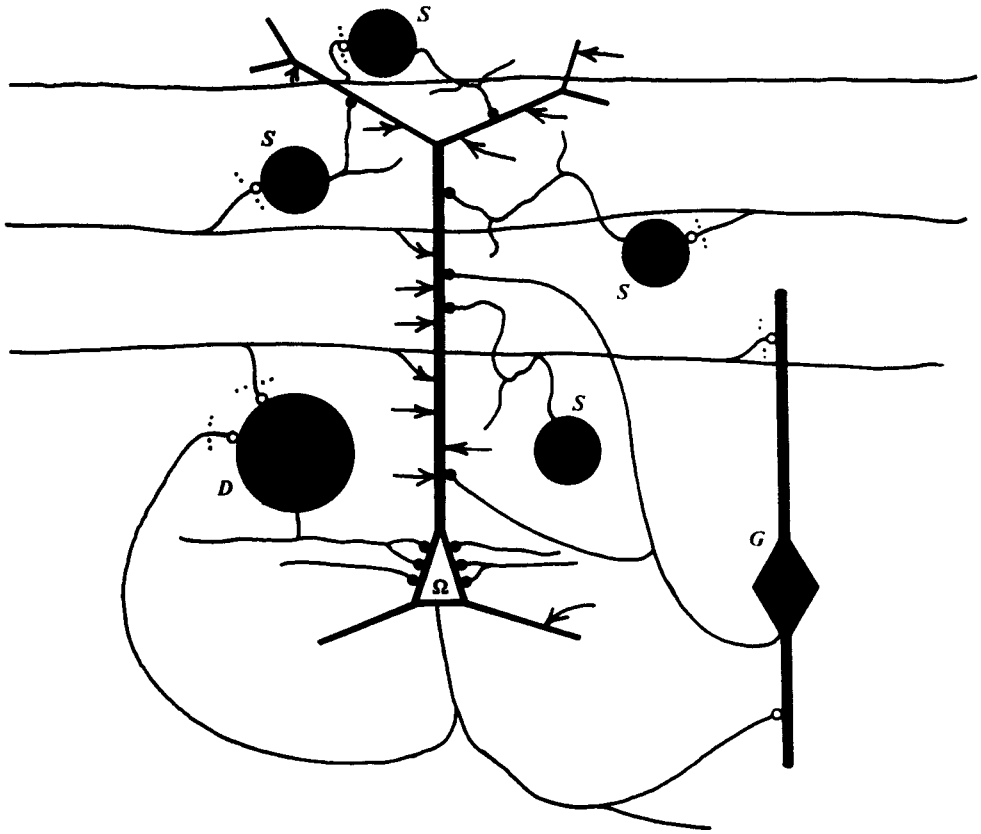

FIGURE 4. The output cell Ω has three kinds of afferent synapse: Brindley synapses (arrows) from codon cells, and two kinds of inhibitory synapses. Those from *S*- and *G*-cells are spread over the dendritic tree (cf. figure 3), and their effect is subtractive: those from the *D*-cells, concentrated at the soma, perform a division.

The problem of the formation of classes for the simple representation of an input has much in common with the problems surrounding codon formation. The central requirement is to choose, from the given population of cells, those which are best suited to representing the current input. This is exactly the problem that was discussed in § 2.1.1, and the possible mechanisms are again those of figure 2. For the same reasons as were given there, Brindley synapses provide the most suitable method of selecting such cells, and may therefore be expected at the cells involved in a simple representation.

It is interesting to note that output cells for the random classes involved in a simple representation require Brindley synapses, whereas output cells for classificatory units proper are best served by having climbing fibres. The freedom allowed by Brindley synapses—independence of different dendrites, and the ability to choose the most appropriate cells—which is such an advantage in simple memory, is only a disadvantage in the neocortical representation of classificatory units. The reason is that in the neocortex, it is crucial that all the relevant evidence for deduction of a property be held at the synapses of a single cell. Modification conditions have to occur everywhere on its dendrites simultaneously, and for all (or enough) of the relevant subevents. Without a climbing fibre, this cannot easily be arranged: a cell which is optimal for one subevent is not especially likely to be optimal for its neighbours as well.

The second part of output cell theory for a simple representation concerns the diagnosis of incoming events. Most of the problems that arise have been considered in output cell theory for the neocortex (Marr 1970, §4.1). These arguments show that two kinds of inhibition are needed: one to perform a subtraction (the *S*-cells of figure 4), and one to perform a division (the *D*-cells or basket cells of figure 4). Such cells would cause the output cells' firing rates to be proportional to $f-p$. In the present case, however, some further information is available: the output cells for a particular event were originally selected (through Brindley synapses) because they had the greatest number of active afferent synapses. Such cells will therefore tend to have more modified active afferent synapses during recall than other cells, and preliminary selection can usefully be made by subjecting the population of output cells to a suitable absolute threshold T (say). In figure 4, it is imagined that inhibition to produce this is provided by the *G*-cells (driven in part by output cell axon collaterals). *G*-cells thus have two functions: to arrange suitable modification conditions during the storage of an event, and to provide a (variable) absolute threshold T during recall. It will be shown in §3.3 that the introduction of two kinds of threshold into output cell theory—i.e. specifying both T and a lower bound on f—greatly improves the performance of a memory.

In figure 5, the apparatus of figure 3 is added to that of figure 4 to produce the basic unit of simple memory. This type of model is examined in detail in §3.

2.2.3. *Structural differences between archicortex and neocortex*

There are various differences in the fine structure of the models devised for archi- and neocortex, of which perhaps the most striking concerns the absence of climbing fibres in archicortex. It is also possible to deduce differences that are predicted by the theory and which concern the large-scale arrangements of the two structures. If all of a large population of output cells tend to receive afferents from the same collection of evidence cells, the disposition of cells and fibres will contrast strongly with their arrangement in neocortex, where one expects that evidence cells are relatively private constructions. There is no reason in archicortex to have evidence and output cells particularly near one another: one can therefore expect to find cells involved in different stages placed rather far apart, and joined by powerful projections. (The so-called perforant path in the hippocampal formation may be an example of such a projection.)

For this reason, the numerical analysis which follows (§3.1) deals with layers of cells $\mathscr{P}_i$, which project to one another with various contact probabilities. Some layers will contain evidence cells, and some, output cells. The difference is however unimportant except in calculations about the recalling abilities of the system.

2.3. *The basic equation, and various constraints*

The calculation of the capacity and recalling ability of the simple memory described in § 2.2 rests on various assumptions and approximations. These are set out together in this section, and the relations derived here are used in § 3.

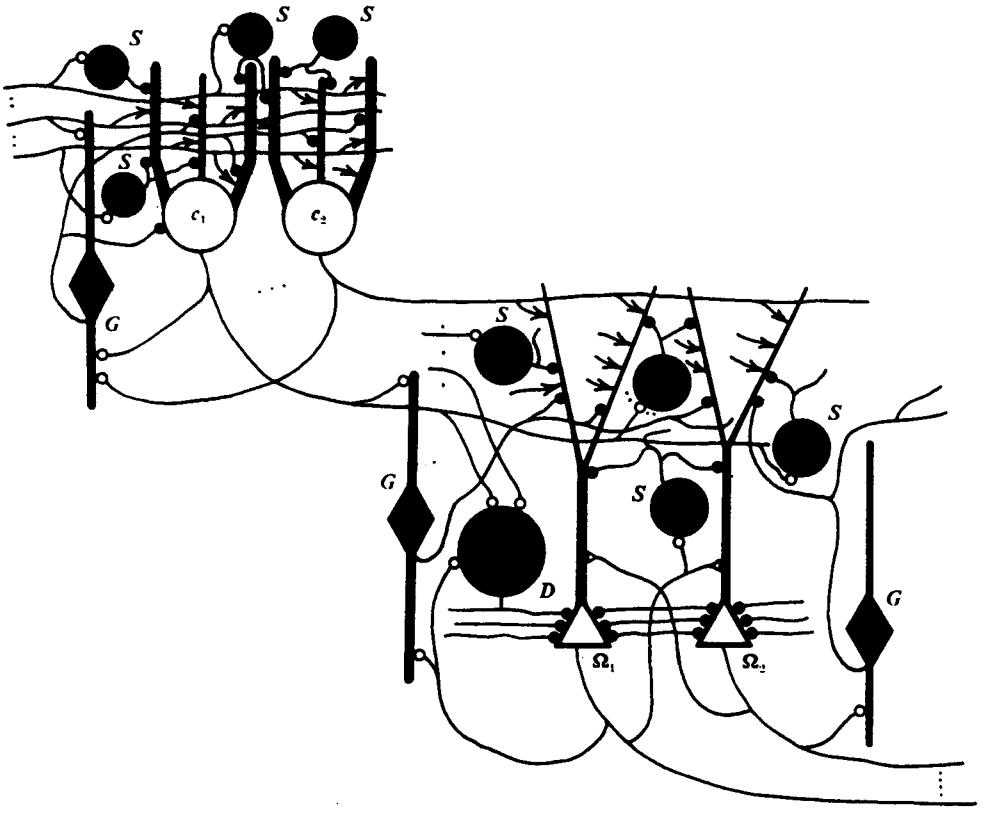

FIGURE 5. A model for simple memory, obtained by combining figures 3 and 4. The output cell axons return to the cells of the current internal description, after giving off collaterals which terminate in Hebb synapses at other output cells. This kind of model is analysed in § 3.1.

2.3.0. *Notation*

$\mathscr{P}_i$ ($i = 0, 1, 2, \ldots$) is a population of N_i cells with activity α_i. The set of cells of $\mathscr{P}_i$ which fire in response to an input is called the *$\mathscr{P}_i$-representation* of the input. The terms *event*, *subevent*, and *codon* will have their usual meanings. In addition, the following notation will be standard:

- E denotes an expectation;
- N_i the number of cells in $\mathscr{P}_i$;
- L_i the number of active cells of $\mathscr{P}_i$ $(L_i = \alpha_i N_i)$;
- R_i the threshold of the cells in $\mathscr{P}_i$ during the storage of information;
- S_i the number of afferent synapses possessed by each cell of $\mathscr{P}_i$ (assumed constant over $\mathscr{P}_i$);
- Z_i the contact probability for the projection of the afferent fibres to $\mathscr{P}_i$ (usually from $\mathscr{P}_{i-1}$). (Thus Z_i = the probability that an arbitrary cell of $\mathscr{P}_i$ receives a synapse from an arbitrary cell of $\mathscr{P}_{i-1}$);
- Π_i the probability that an arbitrary afferent modifiable synapse in $\mathscr{P}_i$ has been modified.

2.3.1. *The response in $\mathscr{P}_i$ to an input event*

If it is assumed that the afferents to $\mathscr{P}_i$ distribute there randomly with contact probability Z_i, the variables defined in § 2.3.0 are related by the following equation:

$$E(L_i) = N_i \sum_{r=R_i}^{L_{i-1}} \binom{L_{i-1}}{r} Z_i^r (1-Z_i)^{L_{i-1}-r} \quad \text{(Marr 1970, § 3).} \tag{2.1}$$

L_i is the sum of expectations (corresponding to the individual terms of the expression), of which one (obtained by putting $r = R_i$) will usually be far larger than the rest. This is because R_i will usually be chosen to keep α_i rather small, which implies that only the terms in the tail of the binomial distribution are in practice used.

2.3.2. *Modifiable synapses in $\mathscr{P}_i$*

It is helpful to have a rough guide as to when it is useful to have synaptic modification at the cells of $\mathscr{P}_i$. Fortunately, it is easy to obtain a simple approximate criterion for this. $\alpha_i = L_i/N_i$ is the activity in $\mathscr{P}_i$: let $\alpha_{i-1} = L_{i-1}/N_{i-1}$ be the activity of the input fibres. This is done because the input to $\mathscr{P}_i$ will be from the population of cells $\mathscr{P}_{i-1}$. It is roughly true that the proportion of synapses active at each active cell of $\mathscr{P}_i$ is α_{i-1}: it is certainly at least this; the amount by which it exceeds it decreases as the value of $S_i\alpha_{i-1}$ increases. Therefore, the probability that after n events, an arbitrary synapse of $\mathscr{P}_i$ has been facilitated is $(1-\alpha_{i-1})^{n\alpha_i}$, which is approximately $1-\exp(-n\alpha_{i-1}\alpha_i)$ if α_{i-1} is small. It is only worth having modifiable synapses in $\mathscr{P}_i$ if, when the inputs have all been learned, not all the synapses there have almost certainly been facilitated—that is, if $n\alpha_{i-1}\alpha_i$ is of the order of 1. Hence a rough, necessary condition that it be useful to have modifiable synapses in $\mathscr{P}_i$ is

$$n\alpha_{i-1}\alpha_i \lesssim 1. \tag{2.2}$$

2.3.3. *The condition for full representation*

The second constraint also embodies a necessary condition—that the activity in $\mathscr{P}_i$ provides an adequate representation of the input event. In the present context, a rather weak criterion of adequacy is sufficient, namely that a change in the firing of the input fibres should produce a change in the cells which are firing in $\mathscr{P}_i$.

The probability that an arbitrary but fixed active input fibre to $\mathscr{P}_i$ does not terminate at any active cell of $\mathscr{P}_i$ is approximately $(1-S_i\alpha_{i-1}/L_{i-1})^{L_i}$. This is approximately

$$\exp(-\alpha_{i-1}S_iL_i/L_{i-1}) = \exp(-S_i\alpha_iN_i/N_{i-1}).$$

Most of the active cells of $\mathscr{P}_i$ would cease to fire if one of their active afferents were removed (by the remarks of § 2.3.1 about the tail of a binomial distribution), and hence the condition for full representation of the input in $\mathscr{P}_i$ is that the probability $\exp(-S_i\alpha_iN_i/N_{i-1})$ should be kept very small—say less than e^{-20}. The condition then becomes

$$S_i\alpha_iN_i \geqslant 20N_{i-1}; \quad \text{i.e.} \quad S_iL_i \geqslant 20N_{i-1}. \tag{2.3}$$

If $\mathscr{P}_i$ is being used to capacity, i.e. $n\alpha_{i-1}\alpha_i \sim 1$, we find that

$$S_iN_i \gtrsim 20L_{i-1}n. \tag{2.4}$$

2.3.4. *Four practical constraints*

It must always be remembered that the cells and synapses of $\mathscr{P}_i$ are physiological objects, which cannot be asked to perform unrealistic feats. One tendency of the theory is to use the populations $\mathscr{P}_i$ of cells with very low activities, α_i. The thresholds of the cells in $\mathscr{P}_i$ have, however, to be set by negative feedback devices, like the *G*-cells of figure 3, and these are to a certain extent limited as to what they can do.

The basic difficulty lies in specifying the proportion of active afferent synapses to which a cell may reasonably expect to be sensitive. Negative feedback devices like the *G*-cell will operate by measuring afferent synaptic activity, and inhibiting the cells with which they synapse in such a way as to keep α at the appropriate value. In what follows, α will be assumed to exceed 0.001 since this figure seems about as small a fraction of active synapses as would allow the activity to be reliably detected. The true bound may be lower, but it cannot be a great deal lower, and certainly not by an order of magnitude.

The same problem applies to the cells of $\mathscr{P}_i$ as applies to the *G*-cells which set their thresholds. In the case where the $\mathscr{P}_i$-cells have Brindley modifiable afferent synapses, the conditions on $\mathscr{P}_i$-cells are probably more stringent than on their associated threshold controllers, since it seems plausible that a considerable degree of post-synaptic depolarization is necessary in a region of dendrite before the conditions for modification are created there. It is difficult to give a numerical translation of the condition on the proportion of active synapses necessary for implementing modification conditions: in what follows, the relevant lower bound will be taken to be 0.005. In practice, it will be possible to alleviate this difficulty by arranging for related synapses to be placed near one another on a dendrite.

Finally, the second tendency of the theory is to require that the number of synapses on a cell be as large as is plausible. Cragg (1967) has shown that the average number of synapses per cell in monkey motor cortex is 60000, and in monkey striate cortex it is 5600. Large archicortical cells are comparable with large motor pyramidal cells, so it is wise to restrict the possible value of S_i to not much more than 60000. An absolute bound of $S_i \leqslant 10^5$ will always be assumed.

There is no direct information about the numbers of synapses on archicortical cells, or the contact probabilities of the various projections, or the activities (α_i) of the various groups of cells. It will not be possible to apply detailed quantitative tests to the present theory's predictions until numerical information of this kind becomes available.

2.4. *The collateral effect*

Let $\mathscr{P}$ be the population of cells in which the simple representation of an input is formed. If each cell has about 60000 afferent synapses, then each one can probably learn about 100 input events (cf. the cerebellar Purkinje cells, Marr 1969). Hence, if the population as a whole is to learn about 10^5 events, the activity α of $\mathscr{P}$ must be about 10^{-3}.

Equation (2.2) of 2.3.2 shows that for learning to be profitable in $\mathscr{P}_i$ driven by cells of $\mathscr{P}_{i-1}$, it is necessary that $n\alpha_{i-1}\alpha_i \lesssim 1$. Let $\mathscr{P}_{i-1} = \mathscr{P}_i = \mathscr{P}$: then the condition becomes $n\alpha^2 \lesssim 1$, and is satisfied by the values of n ($\approx 10^5$) and α ($\approx 10^{-3}$) appropriate to the cells of a simple representation. In other words, it is possible to make good use of learning in synapses from the cells of $\mathscr{P}$ to the cells of $\mathscr{P}$—that is, in synapses at cells of $\mathscr{P}$ driven by collaterals of other cells of $\mathscr{P}$. The practical importance of this is that an input to $\mathscr{P}$ need not be sufficient on its own to re-stimulate all the cells of the particular simple representation which that input is designed to stimulate: collateral activity in $\mathscr{P}$ will help the recall process. Provided that the afferent information causes

more than a critical fraction of the active cells in $\mathscr{P}$ to be cells of the required representation, the collateral system will take over, suppress the cells which should not be active, and stimulate those which should. The completion of a partially specified simple representation by $\mathscr{P}$-cell collaterals is called the *collateral effect.* It will be shown that the collateral effect is probably capable of completing a simple representation when the fraction of currently active cells which are in that representation is as low as one third.

The details of the structure required for the collateral effect are as follows:

(i) collaterals distributing in $\mathscr{P}$ with the appropriate contact probability (see § 3);
(ii) Hebb (or Brindley) modifiable synapses where the collaterals meet other cells of $\mathscr{P}$;
(iii) the usual inhibitory threshold controlling cells.

3. Capacity calculations

3.0. *Introduction*

For practical application of the theory, it is essential to have a firm grasp of the kind of performance that may be expected from the basic simple memory of § 2. This section gives the reader direct experience of the available storage and recall capacity, for reasonable values of the important parameters.

Storage of an event will be said to have been achieved when its simple representation has been formed; and recall of that event, when its simple representation has been recovered.

3.1. *Establishing and recovering a simple representation*

There are various arguments which roughly decide the number of cells and synapses in the different portions of the memory that is analysed here. The conclusions are stated first, in the form of specifications of properties of a network which will form simple representations. These conclusions are followed by the arguments which lead to them, and these, by remarks about the memory's storage and recall performance.

3.1.1. *The basic memory*

There are three populations of cells, $\mathscr{P}_1$, $\mathscr{P}_2$ and $\mathscr{P}_3$. The cells of $\mathscr{P}_1$ send axons to $\mathscr{P}_2$, and those of $\mathscr{P}_2$ send axons to $\mathscr{P}_3$. $\mathscr{P}_3$ possesses a collateral system, and it is in $\mathscr{P}_3$ that simple representations are formed. Table 1 shows the basic parameters for each of the $\mathscr{P}_i$, using the notation defined in § 2.3.0. It is imagined that the 10^6 cells of $\mathscr{P}_1$ are split into 25 so-called *blocks* of cells, each of which projects exclusively to a corresponding block in $\mathscr{P}_2$ (see figure 6). The parameters for each block are given in table 2. The projection from $\mathscr{P}_2$ to $\mathscr{P}_3$ has no block structure, and table 3 describes the parameters for this projection. $\mathscr{P}_3$ also possesses a collateral system, which may be regarded as a projection from $\mathscr{P}_3$ to $\mathscr{P}_{3'}$. The parameters for the collaterals appear in table 3 in the column for $i = 3'$. These values have all been obtained using the equations of § 2.3.

The probability that an arbitrary synapse has been modified can easily be calculated if it is assumed that synapses are effectively chosen randomly each time an event is stored. The assumptions behind this have been set out already (Marr 1969, § 5) in the calculation of the capacity of a cerebellar Purkinje cell. Suppose n events have been stored; then the probability Π_i that an arbitrary modifiable synapse in $\mathscr{P}_1$ will have been facilitated is

$$\Pi_i = 1 - (1 - x_i/S_i)^{n\alpha_i},$$

where α_i, S_i are as in §2.3.0, and x_i is the expected number of synapses used at an active cell for one event. x_i is near to R_i, the threshold of such a cell: in fact

$$x_i = \sum_{R \geq R_i} P_i(R) . R,$$

where $P_i(R)$ is the probability that an active cell of $\mathscr{P}_i$ has exactly R active afferent synapses. $P_i(R)$ is calculated from the terms of the equation in §2.3. Table 4 shows values of Π_i for $n = 5 \times 10^4$, and $n = 10^5$ stored events.

TABLE 1. GROSS PARAMETERS FOR A SIMPLE MEMORY $\mathscr{P}_1 \rightarrow \mathscr{P}_2 \rightarrow \mathscr{P}_3$

Cells of $\mathscr{P}_2$ and $\mathscr{P}_3$ possess Brindley modifiable afferent synapses

i ...	1	2	3
N_i	1.25×10^6	500000	100000
L_i	2500	3025	217
α_i	0.002	0.006	0.002

TABLE 2. $\mathscr{P}_1$ AND $\mathscr{P}_2$ OF TABLE 1 ARE SPLIT INTO 25 BLOCKS, EACH HAVING THE FOLLOWING SPECIFICATIONS:

i ...	1	2
N_i	50000	20000
L_i	100	121
R_i	—	31
S_i	—	10000
α_i	0.002	0.006
Z_i	—	0.2

TABLE 3. THE PROJECTION $\mathscr{P}_2 \rightarrow \mathscr{P}_3$ HAS NO BLOCK STRUCTURE, AND HAS THE FOLLOWING PARAMETERS:

i ...	2	3	3′
N_i	500000	100000	100000
L_i	3025	217	200
R_i	—	351	—
S_i	—	50000	10000
α_i	0.006	0.002	0.002
Z_i	—	0.1	0.1

The column $i = 3'$ gives the parameters for the collateral system in $\mathscr{P}_3$.

The expected number of active afferent collateral synapses at a cell of $\mathscr{P}_3$ is 21.7, but has been taken to be 20 for simplicity.

TABLE 4. MODIFICATION PROBABILITIES Π_i FOR MODIFIABLE SYNAPSES IN EACH $\mathscr{P}_i (i = 2, 3, 3')$ AFTER n EVENTS HAVE BEEN STORED

$i = 3'$ gives values for the collaterals in $\mathscr{P}_3$

n	Π_2	Π_3	$\Pi_{3'}$
5×10^4	0.621	0.538	0.181
10^5	0.857	0.787	0.330

3.1.2. *The collateral effect in* $\mathscr{P}_3$

The collateral system in $\mathscr{P}_3$ can aid the recovery of a simple representation in the following way. Suppose that an input X is presented at $\mathscr{P}_1$, and that X is a subevent of a previously learnt event E_0. Let $\mathscr{P}_{30}$ denote the simple representation of E_0 in $\mathscr{P}_3$ and let $\mathscr{P}_{31}$ denote the rest of $\mathscr{P}_3$.

Suppose that X causes firing of C_0 cells in $\mathscr{P}_{30}$, and C_1 cells in $\mathscr{P}_{31}$. Since E_0 has already been learnt, all collateral synapses between cells of its simple representation will have been facilitated. Hence collateral synapses between cells of $\mathscr{P}_{30}$ will all have been facilitated, whereas those between other cells will have no more than the usual probability of having been facilitated.

In order to analyse the effects of the $\mathscr{P}_3$ collaterals, it is assumed that once firing in the collection $\mathscr{P}_3$ has been established by the afferents from $\mathscr{P}_2$, these afferents become silent, and the cells in $\mathscr{P}_3$ are driven solely by the collaterals. The effects of the collaterals alone can be discovered by regarding $\mathscr{P}_3$ as projecting to an identical set of cells, called $\mathscr{P}_{3'}$, in the same way as the collaterals distribute among the cells of $\mathscr{P}_3$. The behaviour of $\mathscr{P}_{3'}$, which represents the new state of $\mathscr{P}_3$ after one 'application' of the transformation on the $\mathscr{P}_3$ firing pattern induced by the collaterals, can then be calculated using the equations of §2.3.

In the present theory, the important question is whether or not the collateral effect can lead to the recovery of the simple representation of E_0. Whether this happens depends on the parameters associated with the collateral distribution, and on the relative sizes of C_0 and C_1. For fixed parameters there is a threshold for the ratio $C_0 : C_1$ above which the collaterals will tend to increase this ratio, and below which they will tend to decrease it. The threshold is of a statistical nature, because above it, the collaterals are more likely to increase the ratio, and below it, they are more likely to decrease it. One has to move a little way away from this threshold before the outcome either way is virtually certain.

The *statistical threshold* (for $C_0 + C_1 = L_3$) is defined as the value of the ratio $C_0 : C_1$ such that the expected effect of the collaterals is to maintain it. It may be calculated as follows.

Let $\boldsymbol{b}$ be an arbitrary cell of $\mathscr{P}_{3'}$, the copy of $\mathscr{P}_3$ to which the collaterals are imagined to project. The number of active afferent synapses at $\boldsymbol{b}$ comes from a binomial distribution $b(L_3; Z_{3'})$ with expectation $L_3 Z_{3'}$ from population L_3. L_3 is the number of active cells in P_3 and $Z_{3'}$ is the collateral contact probability. Hence the probability that $\boldsymbol{b}$ has exactly x active afferent synapses is

$$P_{3'}(x) = \binom{L_3}{x} Z_{3'}^x (1 - Z_{3'})^{L_3 - x}. \tag{3.1}$$

If $\boldsymbol{b}$ is not in $\mathscr{P}_{30}$, the simple representation of E_0, the number of these active synapses that will have been facilitated is drawn from the binomial distribution $b(x : \Pi_{3'})$ with expectation $x\Pi_{3'}$ from population of size x (from the definition (§2.3.0) of Π). Hence if $Q_{3'1}(r)$ denotes the probability that exactly r of the x active afferent synapses to $\boldsymbol{b}$ have been modified,

$$Q_{3'1}(r) = \binom{x}{r} \Pi_{3'}^r (1 - \Pi_{3'})^{x-r}. \tag{3.2}$$

If $\boldsymbol{b}$ is in $\mathscr{P}_{30}$, all afferent synapses from other cells in $\mathscr{P}_{30}$ will have been modified. Hence the number of active afferent modified synapses at a cell in $\mathscr{P}_{30}$ is composed of two contributions: one, with distribution $b(C_0; Z_{3'})$ from cells of $\mathscr{P}_{30}$ with probability $Z_{3'}$, all of which have been modified: and one with distribution $b(C_1; Z_{3'})$ from $\mathscr{P}_{31}$ which have only chances given by (3.2) of having been modified. For the purposes of calculation, this situation has been approximated by assuming that, for a cell in the simple representation of E_0 with x active afferent synapses, the number of those synapses which have been facilitated has distribution

$$b(x; (C_0 + C_1 \Pi_{3'})/(C_0 + C_1)).$$

Hence if $Q_{3'0}(r)$ denotes the probability that exactly r of the x active afferent synapses to $\boldsymbol{b}$ have been modified,

$$Q_{3'0}(r) = \binom{x}{r} (C_0 + C_1)^x (C_0 + C_1 \Pi_{3'})^r (1 - C_0 - C_1 \Pi_{3'})^{x-r}. \tag{3.3}$$

Hence, if the cells in $\mathscr{P}_{3'}$ all have a threshold R, the expected number of active cells that are not in the simple representation of E_0 is

$$C_1' = (N_3 - L_3) \sum_{r \geqslant R} \sum_{x=r}^{L_a} P_{3'}(x)\, Q_{3'1}(r), \tag{3.4}$$

and the expected number of active cells in $\mathscr{P}_{30}$ is

$$C_0' = L_3 \sum_{r \geqslant R} \sum_{x=r}^{L_a} P_{3'}(x)\, Q_{3'0}(r). \tag{3.5}$$

Thus, when all cells of $\mathscr{P}_{3'}$ have threshold R, the effect of the collaterals is to transform C_0 and C_1 into new numbers with expectations C_0' and C_1'. Hence the statistical threshold, as defined above, for recovery of the simple representation of E_0 is that ratio $C_0 : C_1$ for which

$$C_0 : C_1 = C_0' : C_1', \text{ subject to } C_0 + C_1 = C_0' + C_1' \approx L_3. \tag{3.6}$$

In practice, however, the cells will not have a uniform threshold, since the theory allows that division can take place as well as subtraction. The effect of division may be incorporated by assuming that a cell only fires if at least a fraction f of its active afferent synapses have been facilitated: f is called the *division threshold* of the cell. The combined effects of a subtractive threshold T and a division threshold f are to give a cell $\boldsymbol{b}$ of $\mathscr{P}_3$, with x active afferent synapses, a threshold $R = R(\boldsymbol{b})$ where

$$R(\boldsymbol{b}) = \max\{T, fx\}.$$

This transforms C_i' of (4) and (5) into C_i^* where

$$C_1^* = (N_3 - L_3) \sum_{r \geqslant \max\{T, fx\}} \sum_{x=r}^{L_a} P_{3'}(x)\, Q_{3'1}(r), \tag{3.7}$$

$$C_0^* = L_3 \sum_{r \geqslant \max\{T, fx\}} \sum_{x=r}^{L_a} P_{3'}(x)\, Q_{3'0}(r). \tag{3.8}$$

The statistical threshold becomes that ratio $C_0 : C_1$ for which

$$C_0 : C_1 = C_0^* : C_1^*, \text{ subject to } C_0 + C_1 = C_0^* + C_1^* \approx L_3, \tag{3.9}$$

the threshold parameters T, f being chosen to minimize C_0^*/C_1^*. The expectations C_0^*, C_1^* have been computed for the relevant parameters, and selected values appear in the tables 5 to 7. Cases $C_0 + C_1 = L_3$ and $C_0 + C_1 = \frac{1}{2}L_3$ have both been calculated, since it is often better to use the smaller values during recall. The case $n = 10^5$ and $C_0 + C_1 = L_3$ resembles table 6 in the same way as table 7 resembles table 5. Various other tables have been computed, and the statistical thresholds obtained for selected values of L_3 and $Z_{3'}$ are given in table 8.

Three points are worth noting about these results. First, $Z_{3'} = 0.2$ gives a statistical threshold about twice as good as that for $Z_{3'} = 0.1$. Secondly, recovery of the whole of the simple representation depends upon suitable juggling of T and f, and is complete after about 3 cycles. f must start low, and increase as the representation is recovered: T must decrease in such a way that the activity in $\mathscr{P}_3$ is kept roughly constant. And thirdly, the overall performance of the collateral effect is impressive (see table 8): recovery of the whole of the simple representation of E_0 is almost certain for values of about $0.1\,L_3$ greater than the statistical threshold value (assuming that $C_0 + C_1$ is constant).

The collateral effect is valuable in any population of cells where $n\alpha^2 \lesssim 1$. This condition may

often be satisfied in and between regions of neocortex, and the effect may be an important means of providing indirect 'associational' aid for the interpretation of sensory inputs (see Marr 1970, §2.4).

TABLE 5. THE COLLATERAL EFFECT IN $\mathscr{P}_3$

$N_3 = 100000$; $L_3 = 200$; $Z_{3'} = 0.1$. 50000 simple representations have been stored.

C_0	C_1	T	f	C_0^*	C_1^*
100	0	3	1.0	200	6
		6	1.0	188	0
80	20	6	0.8	151	15
		6	0.9	119	3
60	40	8	0.6	70	19
		7	0.7	86	26
50	50	7	0.6	70	82
		6	0.7	73	101
40	60	7	0.6	41	82
		6	0.7	41	101

Statistical threshold ~ 50:50.

TABLE 6. THE COLLATERAL EFFECT IN $\mathscr{P}_3$

$N_3 = 100000$; $L_3 = 200$; $Z_{3'} = 0.1$. 100000 simple representations have been stored.

C_0	C_1	T	f	C_0^*	C_1^*
100	0	6	1.0	188	14
		9	1.0	136	1
90	10	10	0.9	89	8
		7	1.0	86	6
80	20	8	0.9	110	77
		9	0.9	86	27
60	40	10	0.7	38	80
		9	0.8	48	72

Statistical threshold ~ 85:15.

TABLE 7. THE COLLATERAL EFFECT IN $\mathscr{P}_3$

$N_3 = 100000$; $L_3 = 200$; $Z_{3'} = 0.1$. 50000 simple representations have been stored.

C_0	C	T	f	C_0^*	C_1^*
200	0	4	1.0	200	0
		9	1.0	200	0
160	40	4	0.8	167	1
		8	0.8	167	0
120	80	10	0.6	160	9
		11	0.6	148	4
80	120	11	0.4	88	102
		10	0.5	98	61
40	160	8	0.5	24	186
		9	0.5	20	115

Statistical threshold ~ 60:140.

3.1.3. *Recall performance* $\mathscr{P}_2 \rightarrow \mathscr{P}_3$

The analysis of recall performance $\mathscr{P}_2 \rightarrow \mathscr{P}_3$ and $\mathscr{P}_1 \rightarrow \mathscr{P}_2$ follows the same general line as the arguments of §3.1.2, except that the equations apply only to individual blocks. Let E_0' denote the restriction of the input event E_0 to one block β of $\mathscr{P}_1$, and suppose, as in §3.1.2, that E_0 has already

been learnt. A new input event is presented to the block β, A_0 cells of which were active in E_0' and A_1 of which were not. These in turn evoke (in the corresponding block of $\mathscr{P}_2$) B_0 cells which were also active in response to E_0', and B_1 cells which were not. The firing in $\mathscr{P}_2$ causes the firing in $\mathscr{P}_3$ described by the numbers C_0, C_1 of § 3.1.2. The situation when more than one block of $\mathscr{P}_1$ is active can be solved by a simple extension of the methods used for exactly one block. Figure 6 illustrates the recall problem.

TABLE 8. ESTIMATED STATISTICAL THRESHOLDS (s.t.) FOR VARIOUS VALUES OF THE MAIN PARAMETERS

$N_3 = 100000$; $C = C_0 + C_1$; s.t. accurate $\pm 0.05 L_3$.

L_3	Z_3	C	$10^{-4}n$	s.t.
100	0.1	100	5	30:70
	0.1		10	40:60
	0.2		5	15:85
	0.2		10	20:80
200	0.1		5	50:50
	0.1		10	85:15
	0.2		5	30:70
	0.2		10	50:50
	0.1	200	5	60:140
	0.2		5	40:160

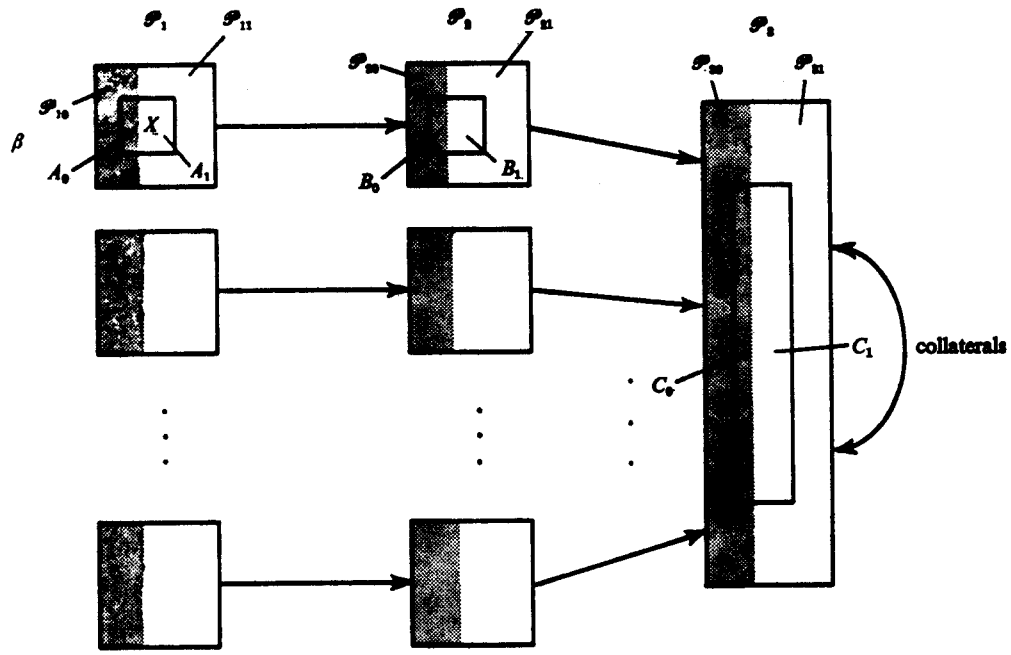

FIGURE 6. The recall problem. $\mathscr{P}_1$, $\mathscr{P}_2$ and $\mathscr{P}_3$ are the populations of cells defined in table 1. Shading represents the parts of these populations involved in the storage of an event E_0. A new subevent X is presented to one block of $\mathscr{P}_1$, A_0 of whose cells were involved in E_0, and A_1 of which were not. This produces activity in one block of $\mathscr{P}_2$, and in $\mathscr{P}_3$. B_0 of the active cells in $\mathscr{P}_2$ were active in E_0, and B_1 were not: C_0 of the active cells in $\mathscr{P}$ were also active in E_0, and C_1 were not. The numbers A_i, B_i, C_i, $(i = 1, 2)$ are computed in the text.

The equations describing the relation between the B_i and the C_j $(i, j = 1, 2)$ are best derived through a series of steps. The notation of § 2.3 is assumed to hold for all processes concerned with the storage of the event E_0; for example, L_3 is the size of the simple representation of E_0 in $\mathscr{P}_3$. The relations between L_i, N_i, R_i, etc., are described by the equations of § 2.3.

S 1. *Additional notation*

The following symbols help to describe states occurring during recall. For $i = 1, 2, 3$:

$\mathscr{P}_{i0}$ = the set of cells of $\mathscr{P}_i$ which were in the $\mathscr{P}_i$-representation of E_0,

$\mathscr{P}_{i1}$ = the set of cells of $\mathscr{P}_i$ which were not in the $\mathscr{P}_i$-representation of E_0.

Thus there are

C_0 cells active in $\mathscr{P}_{30}$,

C_1 cells active in $\mathscr{P}_{31}$,

B_0 cells active in $\mathscr{P}_{30}$,

and B_1 cells active in $\mathscr{P}_{31}$.

Let A_0 be the number of active cells in $\mathscr{P}_{10}$,

and let A_1 be the number of active cells in $\mathscr{P}_{11}$.

S 2. *Calculation of contact probabilities*

The contact probability $\mathscr{P}_2 \to \mathscr{P}_3$ is Z_3, but the contact probability $\mathscr{P}_{20} \to \mathscr{P}_{30}$ is not Z_3, since the cells of $\mathscr{P}_{30}$ were selected (through Brindley synapses) because they had the most active afferent synapses from the $\mathscr{P}_2$-representation of E_0. Let R_3 be the threshold of the cells in $\mathscr{P}_3$ during the setting up of the simple representation of E_0: then the contact probability from the active cells of $\mathscr{P}_2$ to those of $\mathscr{P}_3$ at that time is

$$L_2^{-1}L_3^{-1}\sum_{r\geqslant R_3} N_3\binom{L_2}{r} Z_3^r(1-Z_3)^{L_2-r}r = \xi_0 \quad \text{say:}$$

and the contact probability between active $\mathscr{P}_2$ cells and inactive $\mathscr{P}_3$ cells is depressed slightly: it is in fact ξ_1 where $\xi_1 = (N_3Z_3 - L_3\xi_0)/(N_3 - L_3)$. The contact probability between all other collections in $\mathscr{P}_2$ and $\mathscr{P}_3$ is Z_3. In the following calculations, it will be assumed that distributions between $\mathscr{P}_2$ and $\mathscr{P}_3$ are random, with the contact probabilities ξ_0, ξ_1, Z_3 between the special groups described above.

S 3. *Calculating the number of active synapses at a cell* c *of* $\mathscr{P}_3$

(i) If c is in $\mathscr{P}_{30}$ the number s of synapses active at c is formed from two components: s_0 from the active cells in $\mathscr{P}_{20}$ and s_1 from the active cells in $\mathscr{P}_{21}$. s_0 comes from a binomial distribution $b(B_0; \xi_0)$, and s_1 from a binomial distribution $b(B_1; Z_3)$ (in the usual notation). Hence $P_{30}(s)$, the probability that exactly s synapses are active at c, is

$$P_{30}(s) = \sum_{s_0+s_1=s}\binom{B_0}{s_0}\xi_0^{s_0}(1-\xi_0)^{B_0-s_0}\binom{B_1}{s_1}Z_3^{s_1}(1-Z_3)^{B_1-s_1}.$$

(ii) If c is not in $\mathscr{P}_{30}$ the two components s_0 and s_1 have distributions $b(B_0; \xi_1)$ and $b(B_1; Z_3)$ respectively. Hence $P_{31}(s)$, the probability that exactly s synapses are active at c, is

$$P_{31}(s) = \sum_{s_0+s_1=s}\binom{B_0}{s_0}\xi_1^{s_0}(1-\xi_1)^{B_0-s_0}\binom{B_1}{s_1}Z_3^{s_1}(1-Z_3)^{B_1-s_1}.$$

S 4. *Calculating the number of active facilitated synapses at a cell* c *of* $\mathscr{P}_3$

(i) Let c be in $\mathscr{P}_{30}$ and have s active afferent synapses, made up from the two components s_0 and s_1 of S 3(i). All the s_0 synapses will have been facilitated, and the number of the s_1 synapses

which will have been facilitated has distribution $b(s_1; \Pi_3)$ where Π_3 is the probability that an arbitrary $\mathscr{P}_3$ afferent synapse has been facilitated. So the probability that c has exactly r active afferent facilitated synapses is $Q_{30}(r)$ where

$$Q_{30}(r) = \sum_{s_0=0}^{r} \left\{ \binom{B_0}{s_0} \xi_0^{s_0}(1-\xi_0)^{B_0-s_0} \sum_{s_1 \geqslant r-s_0} \binom{B_1}{s_1} Z_3^{s_1}(1-Z_3)^{B_1-s_1} \binom{s_1}{r-s_0} \Pi_3^{r-s_0}(1-\Pi_3)^{s_0+s_1-r} \right\}.$$

(ii) If c is in $\mathscr{P}_{31}$ the probability $Q_{31}(r)$ that c has exactly r active afferent modified synapses is

$$Q_{31}(r) = \sum_{s \geqslant r} \binom{s}{r} \Pi_3^r(1-\Pi_3)^{s-r}\{P_{31}(s)\},$$

since all active afferent synapses have chance Π_3 of having been facilitated.

S 5. *Calculating the cells' thresholds*

All the cells in $\mathscr{P}_3$ are assumed to be subject to two kinds of threshold: an absolute threshold of T_3 (say), and a division threshold (defined in §3.1.2) of f_3. Thus if a cell has s active afferent synapses, its threshold is set at

$$R_3 = \text{maximum}\{T_3, sf_3\}.$$

S 6. *Calculating expected numbers of active cells*

There are L_3 cells in $\mathscr{P}_{30}$ and $(N_3 - L_3)$ cells in $\mathscr{P}_{31}$. It is assumed that the cells of $\mathscr{P}_3$ are subject to thresholds (T_3, f_3) of S 5. Then the expected numbers of cells active in $\mathscr{P}_{30}$ and $\mathscr{P}_{31}$ are respectively:

$$C_0 = L_3 \sum_{s_0+s_1 \geqslant T_3,\ r \geqslant R_3} \sum Q_{30}(r), \quad \text{where} \quad R_3 = \max\{T_3, (s_0+s_1)f_3\},$$

$$C_1 = (N_3 - L_3) \sum_{s \geqslant T_3,\ r \geqslant R_3} \sum Q_{31}(r), \quad \text{where} \quad R_3 \text{ is as defined in S 5 above.}$$

Close approximations to these distributions have been computed for various values of the important parameters, and some results appear in table 9. They are summarized in §3.1.5.

3.1.4. *Recall performance $\mathscr{P}_1 \to \mathscr{P}_2$*

The problem of describing the effect of presenting a learnt subevent to $\mathscr{P}_2$ can be solved by calculating the values of B_0, B_1 in terms of A_0 and A_1 (defined in S 1 of §3.1.3). These relations are very similar to those holding between the B_i and the C_j $(i,j = 0, 1)$. The following steps S are analogous to those of §3.1.3, and can be derived by the same arguments. Write η_0 for the contact probability between the active cells of $\mathscr{P}_1$ and $\mathscr{P}_2$ during the original setting up, and write η_1 for the contact probability between active $\mathscr{P}_1$ cells and inactive $\mathscr{P}_2$ cells. η_0 corresponds to ξ_0 and η_1 to ξ_1.

S 2 (i) $$\eta_0 = L_1^{-1}L_2^{-1} \sum_{r \geqslant R_3} N_2 \binom{L_1}{r} Z_2^r(1-Z_2)^{L_1-r} r,$$

(ii) $$\eta_1 = (N_2 Z_2 - L_2\eta_0)/(N_2 - L_2).$$

S 3 (i) $$P_{20}(s) = \sum_{s_0+s_1=s} \binom{A_0}{s_0} \eta_0^{s_0}(1-\eta_0)^{A_0-s_0} \binom{A_1}{s_1} Z_2^{s_1}(1-Z_2)^{A_1-s_1},$$

(ii) $$P_{21}(s) = \sum_{s_0+s_1=s} \binom{A_0}{s_0} \eta_1^{s_0}(1-\eta_1)^{A_0-s_0} \binom{A_1}{s_1} Z_2^{s_1}(1-Z_2)^{A_1-s_1}.$$

S 4 (i) $$Q_{20}(r) = \sum_{s_0=0}^{r} \left\{ \binom{A_0}{s_0} \eta_0^{s_0}(1-\eta_0)^{B_0-s_0} \times \sum_{s_1 \geqslant r-s_0} \binom{A_1}{s_1} Z_2^{s_1}(1-Z_2)^{B_1-s_1} \binom{s_1}{r-s_0} \Pi_2^{r-s_0}(1-\Pi_2)^{s_0+s_1-r} \right\},$$

(ii) $$Q_{21}(r) = \sum_{s \geqslant r} \binom{s}{r} \Pi_2^r(1-\Pi_2)^{s-r}\{P_{21}(s)\}.$$

S 5 $$R_2 = \text{maximum}\{T_2, sf_2\}.$$

S 6 (i) $$B_0 = L_2 \sum_{s_0+s_1 \geqslant T_2} \sum_{r \geqslant R_2} Q_{20}(r), \quad \text{where} \quad R_2 = \max\{T_2, (s_0+s_1)f_2\},$$

(ii) $$B_1 = (N_3 - L_3) \sum_{s \geqslant T_2} \sum_{r \geqslant R_2} Q_{21}(r), \quad \text{where } R_2 \text{ is as defined in S 5 of this section.}$$

Close approximations to these distributions have been computed for various values of the important parameters, and selected results are shown in table 10.

TABLE 9. ADDRESSING $\mathscr{P}_3$ WITH AN INPUT, FROM ONE BLOCK OF $\mathscr{P}_2$, WHICH CONTAINS A SUBEVENT OF A LEARNT EVENT E_0

The simple representation of E_0 occupied 217 cells of $\mathscr{P}_3$; n such representations have been stored. Notation is from the text (see figure 6).

B_0	B_1	T_3	f_3	C_0	C_1
$n = 50000$					
120	0	11	1.0	184	27
		12	1.0	166	13
		13	1.0	144	6
		14	1.0	120	3
100	20	13	0.92	101	126
		14	0.92	78	53
		15	0.92	57	21
		11	1.0	56	27
80	40	15	0.75	35	141
		15	0.83	33	79
		13	0.92	51	127
		14	0.92	36	54
60	0	8	1.0	89	110
		9	1.0	58	36
45	15	10	0.75	16	113
		8	1.0	26	110
$n = 100000$					
120	0	17	1.0	53	107
		18	1.0	36	50
100	20	19	0.92	15	144
		17	1.0	23	109
60	0	11	1.0	19	204

3.1.5. *General summary of recall performance*

Table 8 shows the statistical thresholds for recovery of a simple representation in $\mathscr{P}_3$ and tables 9 and 10 can be used to discover the minimal conditions on an input for it eventually to cause the recovery of such a representation. The memory consists of 1.25 million input fibres, divided into 25 blocks of 50000 fibres. A single input event causes activity in 2500 fibres—100 in each block—and the simple representation of each event is formed. Suppose each $\mathscr{P}_3$-cell has 20000 afferent collateral synapses. After 50000 events have been learned, recovery of an event E_0 will have very

high probability of success from stimulation of 30 fibres, all of which were active in E_0, provided that those fibres belong to one block; or from stimulation of 100 fibres in one block, provided that about 70 of those fibres were active in E_0. After 100000 events have been learned, the corresponding figures are 60, and 90 out of 100, still from a single block.

TABLE 10. ADDRESSING ONE BLOCK OF $\mathscr{P}_2$ WITH AN INPUT, FROM ONE BLOCK OF $\mathscr{P}_1$, WHICH CONTAINS A SUBEVENT OF A LEARNT EVENT E_0

The $\mathscr{P}_2$-representation of the part of E_0 in this block occupied 121 cells of $\mathscr{P}_2$; n such events have been stored. Notation is from the text (see figure 6).

A_0	A_1	T_2	f_2	B_0	B_1
$n = 50000$					
20	0	7	1.0	57	50
	0	8	1.0	35	12
30	0	9	1.0	80	26
	0	10	1.0	61	8
40	0	11	1.0	94	12
	0	12	1.0	80	4
80	20	23	0.9	94	5
		24	0.9	89	2
60	40	23	0.8	63	27
		24	0.8	53	14
$n = 100000$					
30	0	11	1.0	43	84
	0	12	1.0	27	27
40	0	13	1.0	64	101
	0	14	1.0	48	39
50	0	16	1.0	67	45
	0	17	1.0	52	18
80	20	28	0.9	73	79
		29	0.9	62	39

3.2. *Justifying the model of* § 3.1

3.2.0. *Generalities*

There are three general constraints which are important in determining the general structure of the memory of § 3.1. They are

(i) that the memory should consist of a number of layers of cells, each receiving connexions from one layer and projecting to one other;

(ii) that the memory needs a capacity, n, of the order of 10^5 events, with good recall capabilities and about 10^6 input fibres;

(iii) that recall should be complete before the projection out of the memory.

The constraint (i) arises because the theory is devised for certain regions of the brain which, according to the available evidence, are connected in this way (see § 4). A theoretician has two general options when designing a memory: he can either specify an exact task, and prove that a particular model is the most economical for that task (cf. Brindley 1969); or he can describe an exact structure, and compute its performance (see, for example, Marr 1969). The present theory has the disadvantage of no exact information; its task is the relating of previously unrelated pieces of knowledge by deduction from plausible general assumptions, the whole being tested by the predictions to which it leads. Condition (i) represents the injection of existing anatomical information into the theory.

Constraint (ii) is important in so far as the design of the memory would have to be changed if

it were shown that the figure of 10^5 was too low. If it were too high, the memory would need only to be shrunk; but a collateral effect is not possible where $n\alpha^2$ is much larger than 1.

It is a matter of common experience that few people can memorize more than 100 randomly chosen items in an hour, though the items may not correspond to the technical term 'event' since many are temporally extended. Even supposing each such item to correspond to 10 events, only 1000 events would need to be stored every hour. This would give 16000 in a 16 h day, which would allow a reasonable number of full days to be accommodated. This seems sufficient for a memory which, it is proposed, is only for temporary storage (information being transferred to the neocortex at least in part during sleep). There is therefore not much danger that 10^5 is an underestimate for n.

The third constraint—that recall should be completed before the return projection—may be justified in two ways. If it is assumed that the return from the memory should occupy as few neocortical synapses as possible, then the return projection must be used only for addressing the neocortical pyramids. There will then be no spare capacity for noise elimination there, and so recall has to be complete before this stage. The second point is that the number of events that may be learned by a single cell is about 100 (§1.2.4). Hence if any neocortical pyramid is likely to be active in such a number of learnt events, all its afferent synapses from the memory will be occupied by the addressing problem. In this case also, there will be no spare redundancy for noise elimination.

These two arguments suggest, but do not compel, the view that the final efferent projection from the memory should perform little more than an addressing task. Constraint (iii) is therefore assumed; but it should be remembered that any spare capacity on the return projection would allow the memory to be correspondingly over-run in its earlier stages.

3.2.1. *The form of the simple representation*

It was shown in §1.4.2 that a model consisting of only one layer of cells (input $\mathscr{P}_1 \rightarrow \mathscr{P}_2 \rightarrow$ return) cannot be constructed to satisfy the general constraints set out in §1. In §3.1, it was shown that a memory with two intermediate layers ($\mathscr{P}_1 \rightarrow \mathscr{P}_2 \rightarrow \mathscr{P}_3 \rightarrow$ return) can. This section discusses how the specifications for $\mathscr{P}_3$ could differ from those of §3.1.

A collateral effect can only be operated usefully among the cells of $\mathscr{P}_3$ if $n\alpha_3^2 \lesssim 1$, i.e. $\alpha_3 \lesssim 0.003$. In order that α_3 be this low, the number of cells in $\mathscr{P}_3$ must exceed 30000, since otherwise the number of active cells in $\mathscr{P}_3$ becomes unrealistically low. N_3 could be say 50000, but the chosen figure was 100000, since this allows a slightly lower α_3 while remaining plausible.

Provided therefore that the need for a collateral effect in $\mathscr{P}_3$ is accepted, N_3 and α_3 must be roughly as in §3.1. If there were no collateral effect in $\mathscr{P}_3$, the constraint that recall has to be complete by then implies that at least one of the projections into $\mathscr{P}_2$ and into $\mathscr{P}_3$ must have low values of Π; i.e. the probability, that an arbitrary modifiable afferent synapse to $\mathscr{P}_2$ or $\mathscr{P}_3$ has been modified, must be low. Hence, either $n\alpha_1\alpha_2 \ll 1$ or $n\alpha_2\alpha_3 \ll 1$. If recall is to be allowed from one block of $\mathscr{P}_2$, Π_3 must be low, and so $n\alpha_2\alpha_3 \ll 1$. Other things being equal, if Π_3 has to be so low that recall is achieved almost totally in $\mathscr{P}_3$ from one block in $\mathscr{P}_2$, α_3 has to be less than it is in the model of §3.1 and thus a collateral effect is possible in $\mathscr{P}_3$.

The arguments are therefore strongly in favour of the form of simple representation shown in §3.1. The memory, if it is anything like that described there, must be rather similar to it. There may of course be other, very different solutions: but the available histological evidence suggests that, for example, the hippocampus is built to a plan along the lines of §3.1 (see §4).

3.2.2. *The specification of* $\mathscr{P}_2$

The block structure in $\mathscr{P}_1$ and $\mathscr{P}_2$ is simply a crude attempt to approximate to an ordering of some kind on the input fibres. The figures chosen have no particular justification: nor does it matter greatly if they are changed.

Once the values of N_3, α_3 have been chosen by the need to create in $\mathscr{P}_3$ a favourable environment for the collateral effect, the shape of $\mathscr{P}_2$ is roughly determined by the number S_3 of synapses allowed for the projection $\mathscr{P}_2$ to $\mathscr{P}_3$. The best use of $\mathscr{P}_3$ requires that Π_3 lies between about 0.2 and 0.8; if α_3 and N_3 are fixed, this roughly determines the number of active afferent fibres that each active cell of $\mathscr{P}_3$ should possess. This determines the relation between L_2 and Z_3, choice of one of these remaining. The final condition, which roughly decides L_2 (and hence Z_3) is the condition that each active afferent to $\mathscr{P}_3$ is received at an active cell of $\mathscr{P}_3$. This fixes an upper bound to L_2 near which (by economy arguments) L_2 should actually be found. The value of L_2 in the model of § 3.1 is 3000, but values up to about 6000 are acceptable, provided slight changes elsewhere are made.

3.2.3. *Input to* $\mathscr{P}_2$

Once L_2 has been roughly decided, the other parameters of $\mathscr{P}_2$ are determined by n (the capacity), and by the input from $\mathscr{P}_1$. For modifiable synapses to be useful in $\mathscr{P}_2$, α_2 must be less than 0.01, and recall performance is much impaired if $\mathscr{P}_2$ does not contain modifiable synapses. This constraint on α_2, together with the rough estimate for L_2, decides N_2. The only remaining numbers are L_1, S_2, Z_2; and the only freedom here is in the choice of S_2, since the conditions (i) $n\alpha_1\alpha_2 \lesssim 1$ and (ii) that L_1 is fully represented in $\mathscr{P}_2$, decide L_1 given S_2. The model of § 3.1 chooses $S_2 = 10000$, giving $L_1 = 100$ per block. $S_2 = 20000$ would allow $L_1 = 200$ per block, but if L_1 is in fact substantially larger than 100, it will be necessary to interpose another layer between the $\mathscr{P}_1$ and the $\mathscr{P}_2$ of § 3.1. (The anatomy of the hippocampal formation suggests that, in the most direct application of this theory, an extra layer of this kind is actually present.)

The general conclusion from the arguments outlined here is that, provided L_1 and N_1 are roughly as in § 3.1, the rest of the memory will have roughly the prescribed dimensions. The specifications of § 3.1 can be changed, and the general equations of § 2 provide rough guides to the consequences of such changes. If L_1 is actually much larger than the value suggested, an extra layer is necessary to transform it into a signal which is acceptable to $\mathscr{P}_2$. Detailed calculations must await the discovery of some quantitative anatomical information.

3.3. *Remarks concerning threshold setting*

3.3.1. *Subtraction and division*

The computations of § 3.1 assumed that inhibition is capable of division and of subtraction. It was proposed by Marr (1970, § 4) that inhibition applied to pyramidal cell dendrites will be subtractive in effect, but that inhibition concentrated at a soma is capable of performing a division. Neither function has been demonstrated to occur.

The model (§ 3.1) does not depend upon the ability to set both a subtraction and a division threshold, but its performance is impaired if only one of these is allowed. If only subtraction is allowed, equations S 5 of §§ 3.1.3 and 3.1.4 become

$$R_i = T_i \quad (i = 3, 2 \text{ respectively}).$$

If only division is allowed, they become

$$R_i = sf_i \quad (i = 3, 2 \text{ respectively}).$$

The equations for the projection $\mathscr{P}_1 \rightarrow \mathscr{P}_2$ have been recomputed for the cases where only a subtraction or only a division function is allowed, and the results appear in table 11. It will be seen that the results, especially for division alone, are much inferior to those when both are allowed.

TABLE 11. COMPARISON OF PERFORMANCE USING PURE SUBTRACTION AND PURE DIVISION THRESHOLDS WITH PERFORMANCE USING A COMBINATION OF THE TWO

Figures are for one block of $\mathscr{P}_1 \rightarrow \mathscr{P}_2$ as in tables 1 and 2. T denotes the subtraction threshold; f, the division threshold. 50000 events have been stored. * denotes no solutions involving between 10 and 1000 active cells. A_i, B_i as in text, and figure 6.

input	subtraction		division		combination	
A_0/A_1	T	B_0/B_1	f	B_0/B_1	(T, f)	B_0/B_1
10/0		*		*	(4, 1.0)	49/354
					(5, 1.0)	23/95
					(6, 1.0)	8/48
30/0	9	80/169		*	(9, 1.0)	80/26
	10	61/48			(10, 1.0)	61/8
	11	43/12			(11, 1.0)	43/2
50/0	13	104/132	1.0	121/393	(13, 1.0)	104/5
	14	94/47			(14, 1.0)	94/2
	15	81/15			(15, 1.0)	81/1
	16	67/5			(16, 1.0)	67/0

3.3.2. *Changing f during recall*

It can be seen from tables 5 to 7 that during the recovery of a simple representation by the collateral effect, best results are obtained if f is raised for each new cycle. In the simple model which was used to make the computations, recovery, if it happens at all, will take place within about three cycles—that is, three successive applications of the collateral effect. In a physiological memory of this type, the cycles as such will not exist in this discrete sense: recovery will be a smooth process. But it will happen quickly, if at all, and will proceed best if f is increased gradually throughout it. The fact that recovery will occur so quickly means that the 'program' for increasing f can without undue loss be the same for all inputs. (This would, for example, not have been so if borderline cases had tended to spend a large number of cycles near the borderline, since f would then sometimes have had to be held for some time at (say) 0.3.)

In physiological terms, this means that the proportion of basket cell inhibition to inhibition applied to the $\mathscr{P}_3$-cell dendrites should initially take some small value—say corresponding to a value $f \approx 0.3$—and should be raised during recall until f is near 1.0. This increase can take place at the same rate and from the same initial value for all recall problems. The likely time-course of the change is of the order of 0.25 s, allowing 50 to 100 ms for each cycle, and the whole operation must be carried out subject to the (negative feedback) condition that a roughly constant number of $\mathscr{P}_3$-cells is kept active. There are various methods by which this could be done, though I can find no single one which seems to be particularly preferable to the others. One method, for example, is to employ an external agency which gradually increases basket cell activity in $\mathscr{P}_3$. The subtractive inhibitory level is then set at an appropriate level by the usual negative feedback through $\mathscr{P}_3$-cell collaterals and an inhibitory interneuron (the G-cells).

3.4. *The return from the memory*

The analysis of the projection back to the neocortical pyramidal cells is straightforward. If, say, each pyramid devotes 10000 synapses to the memory, an expected 22 will be active in each

learnt event. These synapses need to be Hebb modifiable synapses, facilitated by simultaneous pre- and post-synaptic activity. Inhibition needs to be applied to these dendrites so that the cells fire only when all their active afferent return synapses have been modified. In view of the small number active, they probably need to be close together, and perhaps a little larger than other synapses.

3.5. *Scanning during recall*

Simple memory was originally suggested by the need for a direct form of storage which would enable common subevents to be discovered. Addressing the memory with a subevent will cause events to be recalled that contained most of the addressing subevent. Whole events presented to the memory are unlikely to cause recall of other whole events, since any two events will probably differ substantially.

It therefore appears that to use the memory for storage, whole events should be presented to it. Using it for recall requires that subevents should address it, which in turn implies some categorization of the current internal description even at this early stage. The notion that, in order for recall to take place, only a small part of the current internal description should have access to the memory, is close to an idea of attention.

The two problems raised by this are, first, how are common subevents picked out; and secondly, how are they copied out of the memory during the codon formation for new classificatory units? The first problem is the partition problem (Marr 1970, § 1.3.3). Simple memory shows how this problem can be approached, since the ability of a subevent to pick out a related event despite a fair amount of noise shows that test subevents do not have to be all that accurately chosen. Rather general, and perhaps innate, techniques for scanning the current internal description will lead to the discovery of many subevent clusters. The scanning process itself may well be subject to neocortical control. The teaching of scanning techniques—how to 'look' at things—may be a very important factor in the development of a child, since it will have a great influence on the classificatory units that the child will form.

The second problem is more technical and easier to give some kind of answer to. Presumably, when a subevent causes recall of a previous event, it is 'marked' in some way— that is associated (in the technical sense) with a 'marker' input from some special centre. This centre also has a measure of the 'importance' to the organism of this kind of information. When a subevent cluster of sufficient size and importance has been formed, this centre will (perhaps during sleep) call the information out from the memory during a period when codon formation is possible. This can be done simply by addressing the memory with the marker event. The markers have to be fairly simple stereotyped inputs, which can be reproduced when required, and which call up (by association) the subevents that they mark. The obvious candidates for 'marker' inputs, in view of the 'importance' parameter necessary for this function, are the rather primitive firing configurations which may perhaps be associated with the subjective experience of a fairly strong emotion.

The problems outlined in this section will form the subject of a later paper.

4. A THEORY OF HIPPOCAMPAL CORTEX

4.0. *Introduction*

In this section is presented the analysis of hippocampal cortex that follows from its interpretation as a region in which the simple representations of many events are formed. The discussion is restricted to the consideration of local properties of the cortex of various parts of the

hippocampal formation, and includes a brief classification of cortical cell types, based on the results of this paper and of Marr (1970). An interpretation of the macroscopic intrinsic and extrinsic connexions of the hippocampal formation will appear in the paper on hippocampal input–output relations.

4.1. *The morphology of the hippocampal formation*

4.1.0. *Gross morphology*

Most of the following description of the structure of the hippocampal formation is derived from information about the mouse (Cajal 1911; Lorente de No 1934) and the rat (Blackstad 1956; White 1959). There is, however, a remarkable uniformity in the structure of the hippocampus in mammals (Lorente de No 1934), so that the divisions made in the mouse are easily recognizable in man. The only important histological difference is in the size of the elements involved: man's hippocampus is larger in every way than that of the mouse. The homology of the afferent and efferent paths in the two species is less good, since the many slight differences in the sizes of the relevant tracts combine to give overall pictures which are considerably different. Those aspects of the present theory which relate only to histology may however be applied to the hippocampal cortex of most mammals.

Blackstad (1956) and White (1959) have recently used morphological information to classify the various regions of the hippocampal region in the rat. Their findings agree closely, and the present paper will usually follow the terminology of Blackstad. According to that author, the hippocampus admits of the following subdivisions:

(1) area entorhinalis	(a.e.)	
(2) parasubiculum		
(3) presubiculum	(pres.)	
(4) area retrosplenialis e		
(5) subiculum	(sub.)	
(6) cornu ammonis	(CA)	(the hippocampus proper),
was divided into	CA 1	
	CA 2	
	CA 3	
	CA 4	by Lorente de No (1934)
(7) fascia dentata	(FD)	

The division is illustrated in figure 7: Blackstad (1956) gives the explicit criteria for distinguishing the borders between the different regions (1) to (7). Regions (6) and (7) are those most characteristic of the hippocampal formation. The subdivision of (6) CA into CA 1 to CA 4 is based on variations in the structure of the hippocampal pyramids. CA 4 is in many ways distinct from the rest of the CA, and it will be discussed separately in § 4.4, together with the FD (7).

4.1.1. *The histology of the cornu ammonis (CA)*

CA is composed principally of a layer of large pyramidal cells, whose axons constitute the efferent tracts from the hippocampus. Many of these cells are extremely large, and their dendritic trees usually span the whole thickness of the CA. They are arranged in a particularly neat row, and it is the bodies of these cells which give the hippocampus its characteristic appearance. Figure 8 illustrates their arrangement in the cortex.

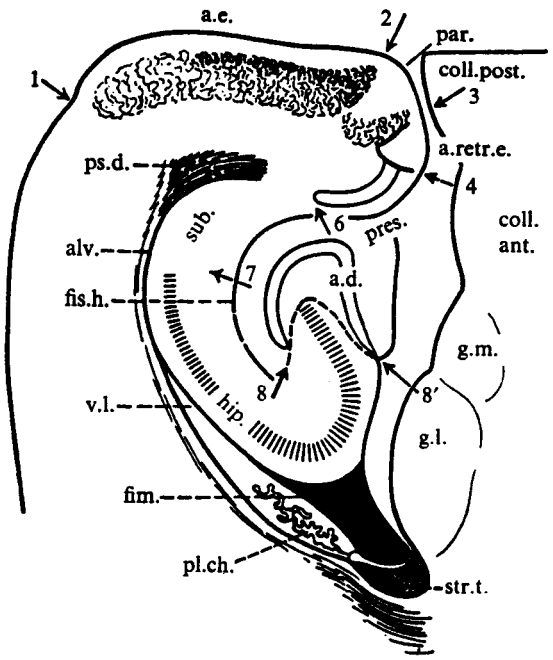

FIGURE 7. Diagram of the hippocampal region in the rat, based on horizontal silver-impregnated sections. The posterior end of the hemisphere is at the top of the figure, the medial side at the right. Arrows show the limits between the areas, which are abbreviated as follows: parasubiculum (par.), presubiculum (pres.), subiculum (sub.), hippocampus (hip.), fascia (area) dentata (a.d.). Other structures shown are ps.d. dorsal psalterium, alv. alveus, fis.h. fissura hippocampi, v.l. lateral ventricle, fim. fimbria, pl.ch. choroid plexus, str.t. stria terminalis, g.l. lateral geniculate body, g.m. medial geniculate body, coll. ant. and post. the anterior and posterior colliculus, and a.retr.e. area retrosplenialis e. (Fig. 2 of Blackstad 1956.)

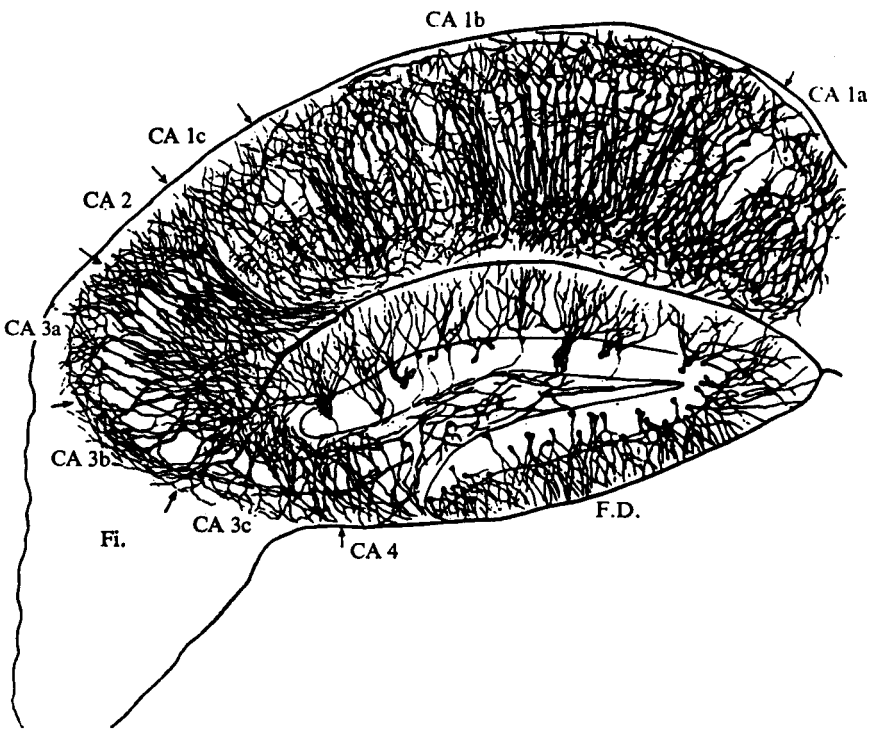

FIGURE 8. Longitudinal section of the adult mouse brain, Cox method. Fi is the fimbria: the divisions are those of Lorente de No (his Fig. 5, 1934).

Hippocampal cortex is commonly regarded as having the four layers shown in figure 9. The bodies of the hippocampal pyramidal cells lie in the Stratum Pyramidale (S. Pyr.), and their basal dendrites span the Stratum Oriens (S. Oriens). Their apical dendrites rise through the Stratum Radiatum (S. Rad.), where they may split into two or more shafts, and arborize freely in the Stratum Moleculare (S. Molec.). The region between the S. Rad. and S. Molec. is often called the Stratum Lacunosum (S. Lac.). Lorente de No (1934) combined information from his own studies with that obtained by Cajal (1911) and earlier authors to give the following description of the cell types in these layers.

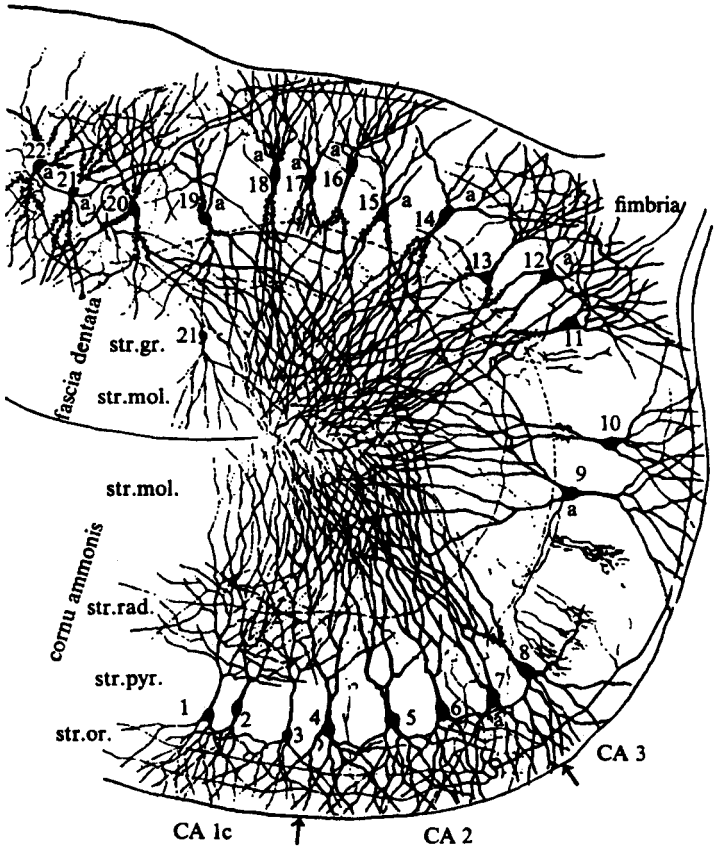

FIGURE 9. Types of pyramids in fields CA 1, CA 2, CA 3, CA 4. 1 to 3 are pyramids of CA 1; 4 to 7 of CA 2; 9 a pyramidal basket cell of CA 3. Only axons of cells 12, 19, 21, 22 have been included in the drawing. Twelve-day-old mouse, Golgi method. (Lorente de No 1934, Fig. 9.)

Stratum Pyramidale

(*a*) *Pyramidal cells.* These vary slightly in appearance from region to region, but figure 9 illustrates their basic uniformity. All pyramidal cells of this class send an axon out of the hippocampus. Those in CA 4 have a modified form, which is explained later.

(*b*) *Pyramidal basket cells.* Their bodies and dendrites are similar to those of the pyramidal cells, but their axons are completely different: they travel horizontally and form baskets round the somas of the pyramidal cells (cell 9, figure 9). There are no basket cells in CA 4, and those in CA 3 do not receive synapses from the so-called mossy fibres (i.e. axons of the granule cells of the FD).

(*c*) *Cells with ascending axon.* Their bodies and descending dendrites are similar to those of the

pyramidal cells, but the ascending dendrites leave the soma directly, and are not branches off a single shaft. The axon arborizes chiefly in S. Rad. (cell 1 of figure 10).

Stratum Oriens

(*d*) *Horizontal cells with ascending axons* have dendrites which remain in S. Oriens: the axons ascend to S. Molec. and arborize there (cell a, figure 12).

(*e*) *Polygonal cells with ascending axons* are similar to (*d*) except in two respects: their axons sometimes emit collaterals in S. Rad., and they send a dendrite to Ss. Rad. and Molec. (cell 5 of figure 11).

FIGURE 10

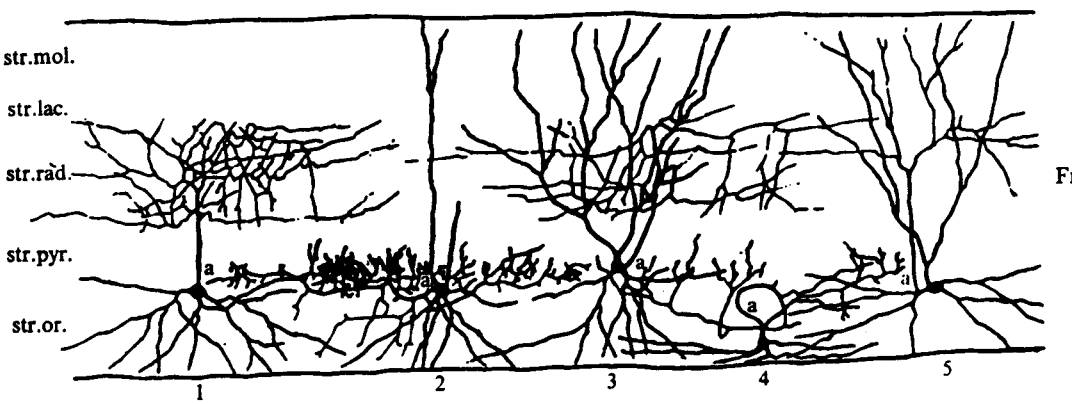

FIGURE 11

FIGURE 10. Types of cell with short axon in CA 1. Twelve-day-old mouse, Golgi method. (Lorente de No 1934, Fig. 7.)
FIGURE 11. Types of cell with short axon in CA 1. Twelve-day-old mouse, Golgi method. (Lorente de No 1934, Fig. 8.)

(*f*), (*g*) *Basket cells* are of two types, one with horizontal dendrites remaining in S. Oriens, and one with a dendrite ascending to S. Molec. (cells 4 of figure 10, 2 of figure 11, and b of figure 12).

(*h*) *Horizontal cells with axon in S. Rad.*, whose dendrites remain in S. Oriens (cell 1 of figure 11).

(*i*) *Horizontal cells with horizontal axon* are globular with dendrites remaining in S. Oriens, and axons ramifying in S. Oriens and occasionally also in S. Pyr. (cells 2 and 5 of figure 10, cell 4 of figure 11).

Strata Radiatum and Lacunosum

Cajal (1911) described the S. Lac. separately in the rabbit, where the Schaffer collaterals are especially distinctly grouped; but in the mouse, cat, dog, monkey and in man, the S. Rad. and S. Lac. are not obviously distinct (Lorente de No 1934). They contain the following types of cell:

(*j*) *Cells with axon ramified in S. Rad.*, of which there are four types, being all combinations of

two kinds of dendritic and two axonal distributions. Some dendrites reach S. Mol., others remain in S. Rad. and S. Lac.; some axons ramify only in S. Rad. and S. Lac., others give branches to S. Pyr. (e.g. cells 3, 6, 7 of figure 10).

(*k*) *Cells with ascending axon ramified in S. Mol.*, after branching in S. Rad. and S. Lac. The dendrites ramify in Ss. Lac., Rad., Pyr. and even Oriens (cells e to m of figure 12).

(*l*) *Horizontal cells of S. Lac.* have axonal and dendritic distributions both in S. Lac., the region of the Schaffer collaterals (see below) (cell 3 of figure 10).

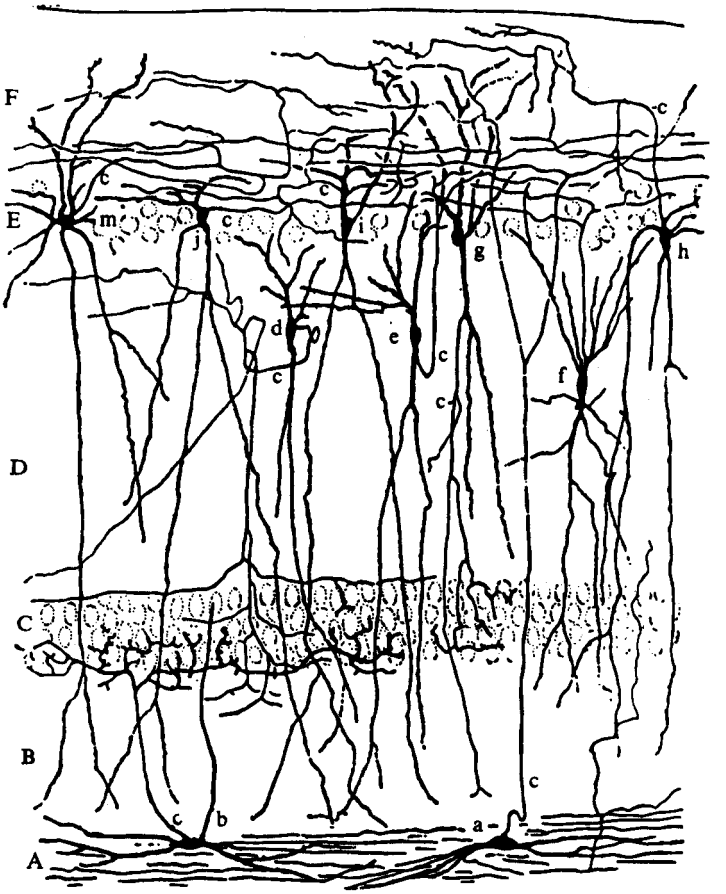

FIGURE 12. Various short-axon cells of the CA. Six-day-old rabbit, double-silver chromate method. (Cajal 1911, Fig. 476.)

Stratum Moleculare

The S. Molec. contains several cells with short axon, typical of a cortical molecular layer.

(*m*) *Cells with short axon*, and

(*n*) *Horizontal cells*,

both of which seem to be rather difficult to stain.

4.1.2. *The histology of the fascia dentata (FD)*

Cajal (1911) gave a full description of FD, which he divided into three layers, the molecular, granular, and polymorph layers. The most notable elements of the cortex are the granule cells, whose bodies, like those of the hippocampal pyramids, are neatly packed and arranged in a granular layer (see figure 13). These cells have supporting cells analogous to those found in CA: they are described on the next page.

Molecular layer

(*a*) *Displaced granule cells* look, and will be treated, like granular cells displaced a little into the molecular layer (cell a, figure 13).

(*b*) *Short-axon cells*, of which there are two main types. The more superficial (figure 14, f and g) have delicate dendrites, mostly horizontal or descending. Their axons are extremely thin and terminate locally, in the outer part of the molecular layer, with a considerable ramification.

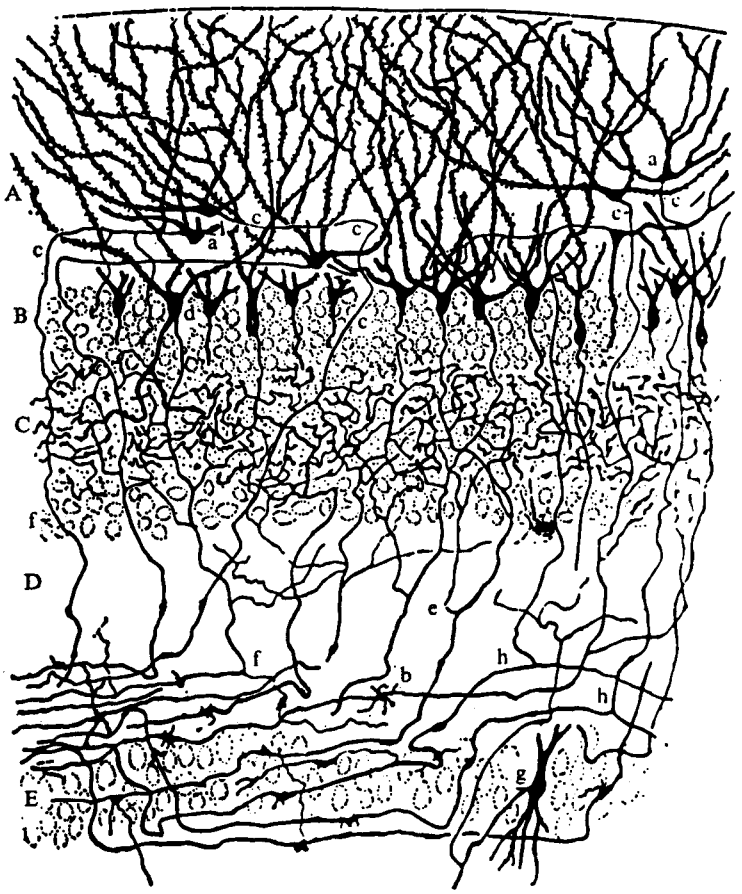

FIGURE 13. The FD in the region of the hilus of the CA. One-month-old guinea-pig, Golgi method. (Cajal 1911, Fig. 478.)

The deeper cells are larger, and occupy the lower portion of the layer (figure 14e). They possess dendrites which spread and divide in all directions—even crossing the granule layer to reach the polymorph layer. Their axons are larger than those of the more superficial cells; they arborize freely in different directions, while remaining in their original layer.

Granular layer

Cajal (1911) regarded the granule cells of the FD as a variant of the cortical pyramidal cells. Figure 13 contains many examples: it will be seen that they lack basilar dendrites, and send about four or five dendrites up through the molecular layer. Their axons are thin, and become the so-called mossy fibres of CA 4 (see below). As they cross the polymorph layer, they give off four or five collaterals, which terminate there. These axons hardly ever give out collaterals after they have crossed the polymorph layer.

Polymorph layer

(*c*) *Pyramidal cells with ascending axon* (figure 15). These cells possess basilar dendrites, which give them a pyramidal shape. Their apical dendrites rise in the manner shown, and their axons eventually ramify horizontally into the granular layer. The cells have obvious similarities with the pyramidal basket cells of the hippocampus proper. Occasionally, but rarely, pyramidal cells are seen that send their axon to the alveus.

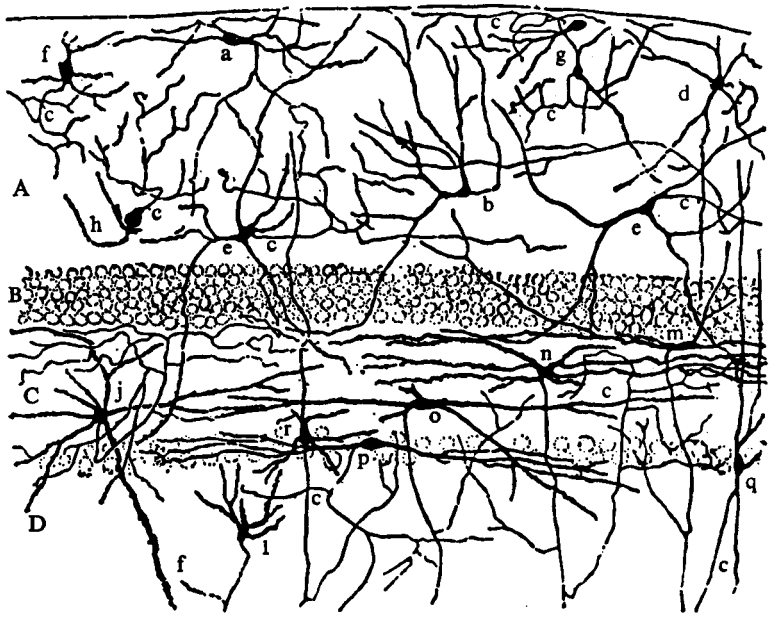

FIGURE 14. The FD. One-month old rabbit, Cox method. (Cajal 1911, Fig. 477.)

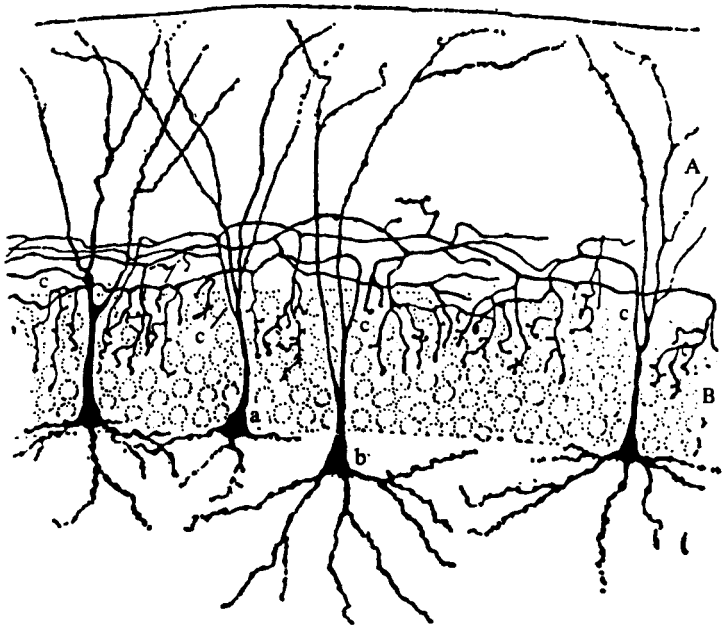

FIGURE 15. The FD. One-month-old rabbit, Cox method. (Cajal 1911, Fig. 480.)

(*d*) *Cells with ascending axon,* which crosses the granular layer and ramifies horizontally. They have various kinds of dendritic distribution (figure 16, e and f; i and o are basket cells).

(*e*) *Cells with descending axon* have long horizontal dendrites which never cross the granular layer. Their axons become fibres in the alveus (figure 16g, j).

(*f*) *Short-axon cells* with local axonal and dendritic distributions: they are found throughout the lower part of this layer (figure 15h).

(*g*) Various star and fusiform cells found low in this layer send their axons eventually to the alveus.

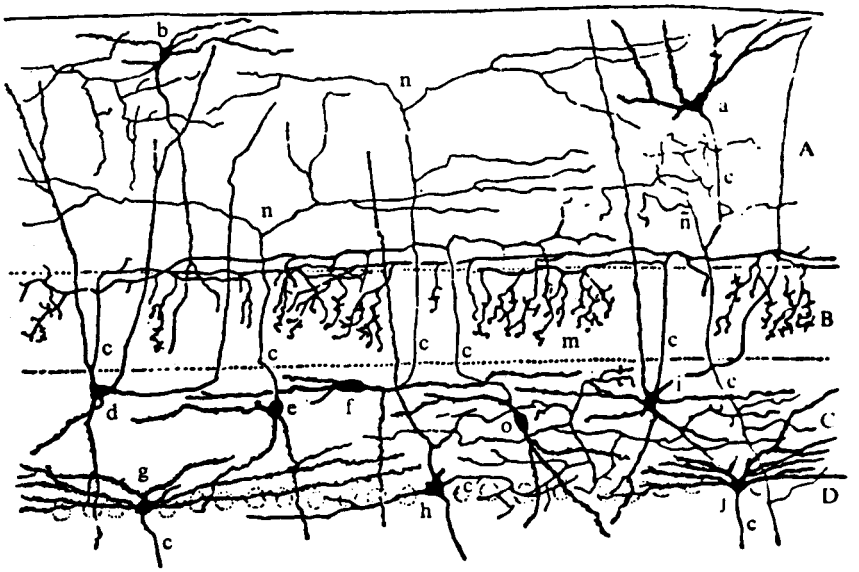

FIGURE 16. The FD. Eight-day-old rabbit, Golgi method. (Cajal 1911, Fig. 481.)

4.1.3. *The principal association systems of the hippocampus*

The present investigation will not concern itself with the relationship between the hippocampi of the two sides of one animal, and consequently little information about the various highly organised commissural connexions will be required (see § 4.5.2). There are four principal systems for association in the hippocampus, and they are dealt with separately.

(i) *The mossy fibres.* The FD granule cell axons become the mossy fibres of the hippocampus. These axons run from FD along CA 4 and CA 3 near the pyramidal cell bodies. They synapse with the dendritic shafts in these regions, producing the distinctive thorns which show up so well in Golgi preparations (figure 9) (Cajal 1911). Few if any penetrate beyond the boundary between CA 3 and CA 2. There are two crucial points to note about these fibres: first, they form the only efferent pathway for the dentate granule cells; and secondly, they specifically avoid the pyramidal basket cells of the hippocampus. These cells thus lack the characteristic thorns (Lorente de No 1934). In CA 4, mossy fibres form the main source of afferent synapses with the pyramidal cells there, and CA 4 contains no basket cells.

(ii) *The Schaffer collaterals* are thick collaterals of the pyramidal cells in CA 3 and CA 4. They travel away from the dentate fascia, and rise through S. Rad. as they go. They synapse in S. Lac. with the pyramidal cells of CA 2 and CA 1 (Schaffer 1892).

(iii) *The axon collaterals of CA* 1 *and CA* 2. The Schaffer collaterals are a transverse association system, joining CA 3 and 4 to CA 1 and 2. CA 1 and 2 also possess a predominantly longitudinal

association system consisting of collaterals synapsing with pyramidal cells in S. Lac.-Molec. These join CA 1 and 2 with other parts of CA 1 and 2. The associations stay more local in CA 1 than in CA 2, but are clear in both cases (Raisman, Cowan & Powell 1965, in the rat).

(iv) *Local associational paths.* It is evident from the descriptions of Cajal (1911) and of Lorente de No (1934) that most hippocampal pyramidal cells have axons which give off collaterals. These probably end locally if they do not contribute to (ii) or (iii), but they have not yet been studied closely. It is necessary therefore to bear in mind that, at least on a local level, the hippocampus is provided with an extremely rich system of interconnexions. It seems to be a general rule in the hippocampus and dentate fascia that different afferent systems terminate both in specific regions and in specific layers of the cortex, not by a random ramification (Blackstad 1956; Raisman *et al.* 1965).

The hippocampal pyramidal cells are extremely large, and so are likely to have at least as many afferent synapses as large pyramidal cells in the motor cortex of the same animal.

4.2. *The hippocampal pyramidal cells*

4.2.0. *The basic model*

The pyramidal cells of sections CA 1, CA 2 and CA 3 of the mammalian hippocampus will be regarded as being populations of cells in which simple representations of various input events are formed. It is proposed that these cells are closely analogous to the cells of $\mathscr{P}_3$ in the model proposed in § 2 and analysed in § 3.

The theory of §§ 2 and 3 requires that, if a cell participates in the simple representations set up in a simple memory of about the specified dimensions, it should have the following properties:

- P 1 Its input fibres should be suitable.
- P 2 The activity α_{CA} of the ammonic pyramids should be small: $0.01 \geqslant \alpha_{CA} \geqslant 0.001$ with α_{CA} probably nearer 0.001.
- P 3 Each cell possesses very many ($\gtrsim 50000$) afferent Brindley synapses from the previous layer of cells, and many ($\gtrsim 10000$) Hebb (or Brindley) synapses from other cells of the CA.
- P 4 Synapses from fibres likely to be co-active should be placed near one another.
- P 5 There should exist an extensive collateral system in CA, giving rise to the collateral synapses of P 3, which allow the completion of the simple representations of partially specified input events.
- P 6 There should exist appropriate supporting cells to supply the required inhibition.
- P 7 There should exist a means of clearing information from these cells when it is re-stored—either as associations or as associations or as new classificatory units—in the neocortex.

Points P 2 to P 7 are discussed separately in the following paragraphs: P 1 is dealt with in a later paper.

4.2.1. α_{CA}

If the hippocampus is involved in storing information in the proposed way, the number of events it can store depends upon the size of each input event, and upon the number of cells used for each. The smaller is α_{CA}, the greater is the capacity, and the more powerful is the collateral effect. α_{CA} is bounded below by about 0.001, a figure which arises out of the necessity to be able to detect those cells which are active (§ 2.3.4). It should not be very difficult to determine α_{CA} by experiment.

4.2.2. *Modifiable synapses*

The competing virtues of the three possible kinds of modifiable synapse of figure 2 have already been discussed. Model 1 was rejected on the ground that each cell would need to be used for more than one input; and the climbing fibre model 3 on the grounds that it needs additional cells, and will not select such suitable cells as model 2 will. It was therefore concluded that model 2, using Brindley synapses, was the preferred choice for all cells in a simple memory. The central feature of Brindley synapses is that they are initially excitatory, and can therefore be used themselves to decide at which cells there should be facilitation (Brindley 1969). This powerful trick solves the problem of selecting the most suitable cells for storing a given input (cf. codon formation, Marr 1970).

There are two practical difficulties associated with the use of Brindley synapses to select CA pyramidal cells for a simple representation. The first arises out of the usual problems associated with a large dendritic tree. It has been pointed out (Marr 1970, § 5.1.4) that in the absence of climbing fibres, it is unreasonable to suppose that synaptic modification is consequent upon simultaneous pre- and post-synaptic activity when these activities are far apart from each other: for example, the spike frequency in an axonal initial segment probably has rather little direct effect upon a synapse 1 mm away at the tip of an apical dendrite of the same cell. Conditions for synaptic modification are therefore likely to hold only locally in a dendrite. This will, however, not be a great disadvantage if input fibres are arranged in such a manner that those that are often coactive tend to lie near one another. It is interesting in this connexion to note that there exists a very marked lamination in the hippocampal afferent system (Blackstad 1956).

The second difficulty is related to the first, and concerns the setting of the thresholds of the CA pyramids. The first time any input is presented to the memory, the appropriate threshold can be computed easily: it is simply a multiple of the power of the unmodifiable component of a Brindley synapse. But after a number of events have been learnt, a non-zero fraction of the CA pyramidal cell afferent synapses will have been facilitated. The thresholds must rise to counteract this effect, and so the amount of inhibition applied to the CA pyramids has to be increased with the number of events that are stored there. Furthermore, if (as seems likely) synaptic modification occurs as a result of a decision process in a local region of dendrite, this inhibition must be applied to such local regions: it is, for example, no use increasing the inhibition at the soma in order to prevent the modification of a synapse at the extremity of an apical dendrite.

The use of Brindley synapses, in output cell selection as well as in codon formation, therefore requires that the amount of inhibition applied to the post-synaptic dendrite, for a given size of input event, should increase with the number of events that the memory has learned. The most satisfactory way of achieving this seems to be to drive the inhibition by collaterals of, in this case, the CA pyramidal cell axons (§ 2 and Marr 1970, § 4.3.1). The cells which achieve this inhibition will be identified in § 4.3.

The conclusion which may be drawn from these arguments, together with those of § 4.3, is that the inhibition level at the CA pyramids can be made to vary in a way which makes it possible for their afferent excitatory synapses to be Brindley synapses. These synapses are in principle the best choice for the function which the present theory assigns to the CA pyramids, and hence the following prediction is made. *Excitatory fibres from the area entorhinalis should terminate on the pyramidal cells of CA* 1 *to CA* 3 *by Brindley synapses.*

4.2.3. *Collateral synapses from other CA pyramids*

The collateral effect (§ 2.4) is an important means by which the simple representation of an incompletely specified input may be completed. The manifestation of this effect in the CA requires that collateral synapses between CA pyramids are modifiable. The synapses included in this discussion are those belonging to reciprocated collateral systems. They do not include either the mossy fibres, or the Schaffer collaterals, both of which are projecting collaterals to which there do not exist reciprocal counterparts: these collateral systems are dealt with in § 4.5.

Collateral synapses should ideally be Hebb synapses: that is, they should initially be ineffective, but should be facilitated by the conjunction of pre- and post-synaptic activity (see § 1.3 for the distinction between Hebb and Brindley synapses). Modification conditions are therefore the same as for the standard CA afferents, except that collateral synapses should probably lack the power to set up modification conditions by themselves.

It is interesting that most collateral synapses to the CA pyramids are found in the S. Rad. (Lorente de No 1934): it seems likely that the importance of the collateral effect is one of the main reasons for the huge development of this part of the dendrite in the CA pyramids. Spencer & Kandel (1961) have shown that the apical dendritic shafts of the CA pyramids can sustain an action potential. It is therefore reasonable to assume that the modification of synapses in S. Rad. could depend on the coincidence of pre-synaptic activity and a burst of post-synaptic action potentials. This would be appropriate on the assumption that decisions about synaptic modification are taken locally in the apical dendritic tree for two reasons. First, spikes will travel at a high rate down through S. Rad. only when that cell is being used to record an input event (though the same activity may lead to the recall of another event): hence post-synaptic depolarization will exist only at the correct times. Secondly, during the recall of an event through the collateral effect, only dendrites in S. Rad. will be exposed to collateral excitation: thus the areas in S. Molec. where the majority of afferents terminate will not be exposed to post-synaptic depolarization, and so inappropriate synaptic modification will not occur there. Both these arguments show that the situation in which the placing of the afferent and collateral synapses was reversed—i.e. where most afferents made synapses in S. Rad.—would be unworkable.

There may be two true reciprocating collateral systems in CA 1 to 3; one distributing its collaterals longitudinally among cells of CA 1 to 2, the relevant fibres rising from S. Oriens and terminating in S. Rad. (Lorente de No 1934); and one being composed of local axon collaterals, many of which distribute in S. Oriens (Lorente de No 1934). Many of the collaterals in the second group will be involved in driving inhibitory threshold controlling cells (§ 4.3). Finally, it must be noted that the associational paths between the hippocampal cortex of each side of the brain must be composed largely of fibres of collateral status. There is evidence that many of these fibres synapse in the contralateral S. Rad. (Lorente de No 1934; Blackstad 1956). (See §§ 4.5.1 and 4.5.2.)

4.2.4. *Numerical predictions*

There are so many unknowns in the equations computed in § 3 that only the most tentative estimates can be made for the expected values of the various parameters. It is probably useful to have some idea of the values compatible with the present form of simple memory theory, if only because if any are shown to be greatly different, it will immediately become clear that others which are related to them must also be different. The following rough values are therefore

given, with the accompanying reservation that they should be regarded only as guides to the orders of magnitude of the various parameters.

(i) α_{CA} is near 0.001.
(ii) $S_{CA} \gtrsim 50000$.
(iii) The number of collateral synapses at a CA pyramidal cell $\gtrsim 10000$.
(iv) Z_{CA}, the contact probability of the afferent fibres, is of the order of 0.1.

4.2.5. *Clearing the simple memory*

The final point with which this section deals concerns the role of the CA pyramidal cells in the transfer of information from simple memory to the neocortex.

The alternative ways of losing information from the simple memory are probably either by a gradual decay applied to all information held therein, or by the selective destruction of a simple representation as the information it represents is transferred to the neocortex. Neither method seems particularly satisfactory: the first would mean that the combination of informations acquired at greatly different times more or less requires that the earlier part has been put into neocortex (a store which, if not actually permanent, is imagined to decay with a rather long half-life). The successful combination probably requires that the earlier has since been rehearsed. The second method is more difficult to make convincing, since the nature of simple memory is such that synapses can be involved in the storage of more than one event: hence the cancelling of one trace has the unwanted side effect of weakening the records of a number of other largely unrelated events.

There seem to be no immediate reasons why either mechanism should be preferred to the other, but the first requires what are probably simpler assumptions about the modification conditions at the hippocampal pyramidal cell synapses.

4.3. *Short-axon cells in the cornu ammonis*

4.3.0. *Introduction*

According to the present theory, the CA contains no codon cells. It follows that none of the short-axon cells found there are excitatory, and that they carry out all the functions required of inhibitory threshold controlling cells. Hippocampal cortex is in this respect unusual: the cerebellar cortex certainly contains short axon excitatory cells (the granule cells, Eccles, Llinas & Sasaki 1966), and the cerebral neocortex probably does (Martinotti cells, Marr 1970).

4.3.1. *The functions of inhibition*

The present theory requires that the thresholds of the CA pyramids be controlled in a very careful manner. Suppose that synaptic modification is an all-or-none process, and that p, q represent respectively the strengths of the unmodified and modified states of a Brindley synapse, where $0 < p < q \leqslant 1$. Then $[p, q]$ is the analogue of the plausibility range for output cells (Marr 1970, §4.1.3).

The three principal tasks of the pyramidal cell threshold-setting mechanisms are as follows:

T 1. *The storage of events:* when an event E is presented, synaptic modification must take place at those cells which have the greatest number of active afferents.

T 2. *The recognition of subevents:* when a subevent X is presented, those cells must fire which have the greatest fraction f of active afferent modified synapses, provided that the number of such synapses exceeds some number, T.

T 3. *The completion of events:* given the firing of a number of hippocampal pyramidal cells, those cells must fire at which the greatest fraction of active afferent collateral synapses have been modified, provided that the number of such synapses exceeds some number T'.

These criteria have to be fulfilled without any other instructions, if possible: that is, the mechanism for performing T 1 should naturally perform T 2 when the current input subevent has occurred in a previous event. Collateral synapses tend to lie in S. Rad., where they have their own special inhibitory cells, so T 3 can to some extent be taken separately. The three tasks are discussed below.

4.3.2. *The storage of events*

The crucial factor in the storage of events is that the correct conditions for synaptic modification prevail in the pyramidal cell dendrites. Excitation there is due to two components: one, of fixed size, due to the unmodifiable excitatory component of the Brindley synapses; and one, whose size increases with the number of events stored in the memory, due to the fraction of active synapses that have already been facilitated.

The first component is a standard multiple of the number of active afferent fibres, and can reasonably be expected to be counteracted by local inhibitory cells in the hippocampal cortex. The function of these cells is to provide inhibition in the pyramidal cell dendrites such that when no events have been learned, only those dendrites which receive more than a certain number of active synapses are depolarized enough to modify their active afferent synapses. (The necessary number of such synapses is the threshold which appears in table 3.) This inhibition can be provided by cells whose axonal and dendritic distributions are subject to the kinds of sampling techniques outlined by Marr (1969). The obvious candidates for such cells in the hippocampal cortex are the components of cells (*c*) and (*e*) due to their ascending dendrites; cells (*i*) (for this function in S. Oriens); (*j*); (*l*); (*m*); and (*n*) (see §4.1.1).

The second component must increase with the number of events stored in the memory, and again must act on the dendrites of the pyramidal cells, where it must affect the formation of postsynaptic conditions for synaptic modification. It was argued in §2 that the simplest way of achieving this is by having inhibitory cells driven by axon collaterals of the hippocampal pyramids (analogous to the upper dendritic tree of the cerebellar Golgi cells). The following cells of §4.1.1 are interpreted as performing this function: the components of cells (*c*) and (*e*) due to their descending dendrites; (*d*); (*h*); and (*k*). This is an important function for which, fortunately, many of the described cells have appropriate axonal distributions. It remains for electron microscope studies to show whether the dendrites of any of these cells receive synapses from the pyramidal cell axon collaterals.

4.3.3. *The recognition of subevents*

It was shown in §§2, 3 that the most sensitive indicator of whether a given cell has previously recorded a subevent similar to the current one is the *fraction* of the active synapses which have been modified. This is computed by a division which Marr (1970, §4.1.6) has argued may be associated with inhibition applied to the soma of a pyramidal cell. The requirement set out in the discussion there of output cell theory was that the amount of inhibition applied to the soma should vary with an estimate of the total number of active fibres: and this is obtained by dendritic sampling by many inhibitory cells, whose synapses converge at the soma. Such cells are for this reason usually

called basket cells, and are present in the hippocampus with suitable axonal distributions (cells (*b*), (*f*) and (*g*) of §4.1.1). Andersen, Eccles & Løyning (1963) have shown that they are inhibitory, but the question of whether they effectively perform a division has not yet been investigated.

The second component of §4.3.2 is also needed for the recognition of learnt subevents.

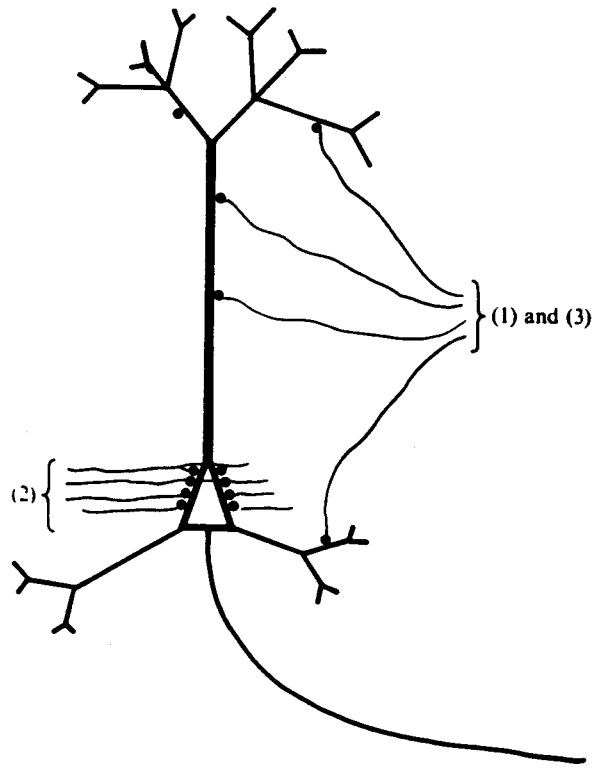

FIGURE 17. Three functions of inhibition: (1) Remove pK where $[p, q]$ is the plausibility range. *S*-cells (i.e. cells *c*, *e*, *i*, *j*, *m*, *n*, of §4.1.1: *l* for the Schaffer collaterals). (2) Divide by K to obtain the fraction f of the active synapses that have been modified. Basket cells (*b*, *f*, *g* of §4.1.1). (3) Raise p to some value p' such that: (*a*) the correct number of cells have outputs in the range $[p', q]$: p' depends on E; (*b*) the correct modification conditions are implemented (cells *c*, *d*, *e*, *h*, *k* driven by pyramid collaterals (§4.1.1)).

4.3.4. *The completion of a simple representation*

According to §2.4, the principal mechanism available for the completion of a subevent X is the collateral effect, which can recover the simple representation of the event $E \vdash X$ even though X is small (§3.1). For this, collaterals of the pyramidal cells should synapse with other pyramidal cells (in S. Rad.) through Hebb (or Brindley) synapses. Recovery of a simple representation by the collateral effect has been discussed at length in §3.1, where it was seen that best results are achieved if the division threshold (basket inhibition) can be gradually increased during recall. The subtractive inhibition must be decreased in a corresponding way, so as to keep the number of active cells roughly constant.

Subtractive inhibition requires inhibitory synapses applied to S. Rad., and for this the cells (*c*) of §4.1.1 would be suitable. Cells (*h*) have the appropriate dendritic and axonal distributions for the division function. Many of the cell types referred to in §4.3.2, however, have axons ramified

in S. Rad. and S. Lac. as well as in S. Molec. This suggests that synapses in S. Rad. and S. Lac. may also be Brindley synapses, and hence that selection of CA cells depends on their suitability judged from the point of view of the collateral effect as well as of the exogenous afferents.

Although there are various ways by which the proportion of somatic to dendritic inhibition might be changed during recall, the available information does not help one to decide if this is in fact done. One possibility is that the transmitter at basket synapses tends to be degraded rather slowly, causing the effect of these synapses to increase gradually during stimulation. The negative feedback circuit through the other cells would ensure that dendritic inhibition is decreased in an appropriate way.

The three functions performed by the inhibitory cells of the CA are summarized in figure 17; the cells thought to be responsible for each are listed in the legend.

4.4. *The fascia dentata*

4.4.0. *Introduction*

The granule cells of the FD will be regarded essentially as extensions to the dendritic trees of the CA pyramidal cells. It is proposed that simple representations are set up in FD in the same way as in CA 1 to CA 3, but that instead of the FD granules sending their own axons elsewhere, they synapse with what may be regarded as 'collector' cells in CA 4 and CA 3. The collector cells send axons elsewhere, and a collateral effect probably operates amongst them.

There are various ideas behind this interpretation of the FD granule cells. The first is that the proposed scheme will result in a saving in the total number of cells transmitting simple representations elsewhere, and hence in savings elsewhere in the numbers of cells and synapses needed to deal with them. It has been seen that the storage capacity for simple representations in a population of cells depends on the activity α of that population; and that α is likely to be bounded below by about 0.001. Hence above a certain point, it is unprofitable to increase the size of the population carrying simple representations, the certain point being in the region of 10^5 cells. If the amount of afferent information to be dealt with requires more cells than this, something like the proposed theory for the FD becomes the natural scheme to adopt.

The second idea concerns α_{FD}, the activity of the FD cells. Once it has become unnecessary for a collateral effect to operate among the cells of a simple representation, the lower bound on α_{FD} ceases to be dictated by the constraint that only about 10000 synapses will be available for the collateral effect. The value of α_{FD} can be pushed down to the bound dictated by the weaker constraints that α_{FD} can be detected by other cells all of whose synapses may be devoted to the task—by the local inhibitory cells, and the proposed collector cells. This notion implies that the collector cells should possess potentially powerful afferent synapses from FD granules, an implication which receives support from the huge size of the mossy fibre synapses in CA. Thirdly, the activity in the population of collector cells must be comparable to that in the rest of CA, so that a collateral effect is possible there.

Finally, it is worth noting that the present theory supports the opinion of Cajal (1911), based on histological evidence, that the dentate granules are a variant of the hippocampal pyramids in CA 3. Lorente de No (1934) remarks (p. 147) that, in the monkey and in man, the similarity between CA and FD is outstanding.

4.4.1. *The FD granule cells*

In the present theory of the FD, essentially the same remarks apply to the granule cells as were made about the CA pyramids, except that there may be no collateral effect amongst them. (It may be replaced by a collateral effect among the cells of CA 4 and CA 3 to which the granules project.) The granule cells (figure 13) are therefore regarded as being like CA pyramids without an S. Rad., S. Lac. or S. Oriens. Their principal afferents from elsewhere should terminate in Brindley synapses: all synapses from local short axon cells should be inhibitory, and should terminate in unmodifiable synapses. The inhibitory synapses on the granule cell dendrites should have a subtractive effect, and those on the soma should perform a division (§4.3 and Marr 1970, §4). The activity α_{FD} should be very small, probably less than α_{CA}. Synapses likely to be coactive should be juxtaposed, and the afferent contact probability is probably in the region of 0.1, and may be greater than that found in the CA.

The present theory gives no grounds for supposing that any granule cells should not possess afferent basket synapses (or an equivalent grouping of inhibitory synapses just above the soma). The special cells noticed by Cajal (cell a, figure 16) are therefore not explained by this theory, unless they are found to be inhibitory and to have a local axonal distribution, or to be extremely rare.

4.4.2. *Short-axon cells in the FD*

The requirements for inhibition in the FD are the same as in the CA 1 to CA 3, and the arguments put forward in §4.3 need not be repeated. It remains only to summarize the different functional elements required in the dentate cortex, and to identify them with the cells described by Cajal (1911). The next three headed paragraphs correspond to the sections 4.3.2 to 4.3.4 on short axon cells in the CA.

The storage of events. It was seen in §4.3.2 that two components of inhibition are required to ensure that the correct numbers of synapses are modified by an incoming event. The first varies only with the number of active afferent fibres, and is performed by short axon cells with local dendritic fields. Such cells estimate the amount of local afferent fibre activity, and send inhibition to the granule cell dendrites (cells *b* of §4.1.2, including only those parts of the activities of cells e of figure 14 that are due to dendrites in the molecular layer). The second component of inhibition must increase with the number of events stored in the memory. It should be supplied by cells whose axons ramify in the molecular layer, but whose dendrites are exposed mainly to activity in granule cell axon collaterals. The polymorph layer, below the granule cell bodies, receives most of their collaterals: the natural candidates for these inhibitory cells are *b* and some of *d* of §4.1.2.

The recognition of subevents requires basket cells and the cells of the last paragraph. Basket cells are present in the FD (cells *c* and others of *d* of §4.1.2).

The completion of subevents relies on the collateral effect. Although it is thought that this principally occurs in CA 4 and CA 3, it is worth noting that some FD granule cells do send axon collaterals to the molecular layer of the FD, where the appropriate inhibitory mechanisms are already available.

Remarks. The only cells left unaccounted for are certain inhabitants of the polymorph layer (cells, *e*, *f*, *g* of §4.1.2). It seems likely that these cells, found principally in the lower parts of the polymorph layer, should properly be regarded as components of CA 4: the long axon cells as 'collectors' (see later) and those with short axon as the usual inhibitory threshold controlling cells.

There is some evidence (Raisman *et al.* 1965) that septal afferents to the FD terminate in the polymorph layer, though this is not firmly established. If it is true, and if the polymorph cells are largely inhibitory, the finding suggests that the septal nuclei play rather a special role in controlling the FD.

4.4.3. *CA 3, CA 4 and the mossy fibres*

Lorente de No (1934) described the large cells of CA 4 as modified pyramidal cells. They differ in two major respects from the pyramids of CA 3: first, no basket plexus envelops their somas; and secondly, they receive mossy fibre synapses over much of their dendrites, not (as in CA 3) over small sections of dendrites near the soma.

Since no basket plexus envelops the somas of the CA 4 modified pyramids, it follows that the mossy fibres fail to drive basket cell inhibition at these cells. This interesting characteristic is preserved by the mossy fibres in CA 3, where they conspicuously avoid synapsing with the pyramidal basket cells. No other hippocampal afferents share this feature.

In that part of CA 4 which is closest to FD, almost the whole of the modified pyramids' dendrites seem to be covered with long spines: the number appears to decrease slightly towards CA 3. At the border between CA 3 and CA 4, two things happen: the pyramids suddenly start sending a dendritic stem to the molecular layer of the CA, so the number of their afferent fibres that are not mossy increases sharply; and the basket plexus appears (Lorente de No 1934).

It was proposed in § 4.4.0 that the cells of CA 4 are essentially collector cells for the FD granules, in which an output representation of FD activity is set up and transmitted elsewhere. Thus if mossy fibre synapses are modifiable, they are Brindley synapses, and the setting up process proceeds in the usual way. For this, inhibition is required in CA 4, so that only the correct, small proportion of CA 4 cells is used each time. Short-axon cells of the required kind have been described by Lorente de No (1934, p. 132). The situation is in outline the same as for the ordinary pyramids of CA 1 to CA 3, and the remarks of § 4.3.1 about the setting-up process apply here.

One of the two anomalies concerning the mossy fibres—that they produce very large synapses (Hamlyn 1962) and are not associated with basket inhibition—can be explained by assuming that α_{FD} is extremely low. For this means that P(CA 4 & FD), the probability that a (randomly chosen) CA 4 pyramid and an FD granule fire simultaneously, is extremely small—less than P(CA 3 & CA 3) for example—and hence that the mossy fibre synapses should be larger than the CA 3 to CA 3 collateral synapses. The fact that the mossy fibres do not drive basket inhibition may mean that these synapses are not modifiable.

4.5. *Collaterals and their synapses in the hippocampus*

4.5.1. *Collaterals in the CA*

All hippocampal pyramidal cells send collaterals to S. Oriens (Lorente de No 1934), of which those from CA 2 seem to be the longest. Most give off ascending collaterals which ramify locally in S. Rad., and many also produce a major long-distance collateral to S. Lac. This last category includes the Schaffer collaterals from CA 3 and 4 to CA 1 and 2, and the longitudinal collaterals which arise from cells in CA 1 and 2, and from those cells of CA 3 which have no Schaffer collaterals (Lorente de No 1934).

The collateral effect proper (§ 2.4) is thought to be associated principally with the local axon

collaterals which ramify in S. Oriens and S. Rad. If S. Oriens and S. Molec. are largely independent (a conjecture suggested by their great distance apart), the collateral effects to which each gives rise could be largely independent. Collaterals in S. Oriens are also expected to drive recurrent inhibition (§§ 4.3.2 and 4.3.3).

The long-distance collaterals probably serve another function, analogous to that proposed for the mossy fibres. The axons of the cells of CA 3 and 4 project in the rat to the septal region only; those of CA 1 and 2 project to the anterior thalamus, the mammillary bodies, and to the septum (Raisman, Cowan & Powell 1966). Thus the cells of CA 3 and 4, and hence also of FD, have access to the mammillary bodies and the anterior thalamic nuclei only through the Schaffer collaterals. It is not known to what extent the CA 1 and 2 longitudinal collateral system is a reciprocal one, so it is not possible to say what kind of collateral effect these fibres produce. The efferent projections from CA 1 and 2 are to a certain extent topographically organized (Raisman *et al.* 1966), so the only way one part of (say) CA 2 can influence cells to which another part projects is probably through the longitudinal association path. Such associational effects may require that the relevant collateral synapses are Hebb (or Brindley) synapses, and that the cortex is supplied with suitable inhibitory interneurons (e.g. cells *l* of § 4.1.1 for the Schaffer collaterals).

The afferent fibre systems to the hippocampus are also to some extent topographically organized (Raisman *et al.* 1965). It is therefore possible that a subevent may be fed into CA 3 and 4 alone: this subevent may previously have been associated with a simultaneous subevent in CA 1 and 2, but this may now be absent. The input to CA 3 and 4 can, through the Schaffer collaterals, evoke the original activity in CA 1 and 2 by stimulating cells there and relying on a local collateral effect (in the usual way). Provided (*a*) that the activities α in CA 1 to 4 are low enough for this simple kind of association to work (in conjunction with a local collateral effect), and (*b*) that the Schaffer collateral synapses are strong enough to allow rather few active facilitated synapses to stimulate a cell in CA 1 and 2, these collaterals could initiate this kind of associative recall. The higher the probability that a given Schaffer collateral synapse has been modified, the higher the number of facilitated collateral synapses that needs to be active at a CA pyramid in order for that cell to fire.

Hamlyn (1962) and Andersen (1966) describe the Schaffer collateral synapses as having a size between that of the usual spine synapses, and that of the mossy fibre synapses. This suggests that the probability that a Schaffer collateral synapse has been modified lies between the values for the other two kinds of synapse: i.e. if the probabilities that an ordinary collateral, a Schaffer collateral, and a mossy fibre synapse have been modified are p_c, p_s, p_m respectively, one would expect that $p_c > p_s > p_m$.

4.5.2. *Commissural connexions*

Blackstad (1956) found that most hippocampal commissural fibres are very fine, and terminate in the Ss. Oriens and Rad., with a certain number from the contralateral area entorhinalis to S. Lac.-Molec. He was unable to determine the origins of many of these fibres, but from his evidence, and that of Raisman *et al.* (1965), it would seem that the projections are probably homotopic in CA 2 to 4, and are certainly homotopic and very symmetrical in CA 1.

The details of these projections are unimportant at the present crude level of theory: it is important only to note that, since the connexions are probably reciprocal, they probably allow a standard collateral effect (§ 2.4) between the hippocampi of the two sides. It is in accordance with

the theory that those fibres which terminate above S. Pyr. do so in S. Rad. rather than in S. Molec.; and with the notion that S. Molec. and S. Oriens are independent that they should distribute both above and below S. Pyr.

4.5.3. *The FD*

Cajal (1911) and Lorente de No (1934) both describe the collaterals of the dentate granule cells. They synapse with the dentate polymorph cells (as required by § 4.4.2), and to some extent they ramify in the molecular layer. This would enable something of the usual collateral effect to take place among the dentate granules.

Blackstad (1956) describes massive degeneration in the inner one-quarter to one-third of the molecular layer after contralateral lesions, but is uncertain of the origin of the fibres responsible. Raisman *et al.* (1965) have some evidence which implicates the contralateral septum, but suspect there may be a projection from the contralateral CA 1.

4.6. *A brief functional classification of cell types*

4.6.0. *Introduction*

The distinction between archi- and neocortex is thought to reflect a difference in their functions. Archicortex is essentially *memorizing cortex*, in the sense that a given area of archicortex is likely to contain one or more layers of a simple memory. It typically contains cells resembling the hippocampal pyramids or the dentate granules, without climbing fibres. Neocortex, on the other hand, though undoubtedly used a great deal for simple associational storage, can probably be regarded as *classifying cortex*. Its operation depends on climbing fibres, and its success depends upon the truth of the fundamental hypothesis (Marr 1970, § 1.6.4).

In the following sections 4.6.1 and 4.6.2 are listed the principal types of cell which the theories predict in memorizing (M) and in classifying (C) cortex. In general, archicortex is memorizing cortex, and neocortex can do both. Special additional considerations probably apply to those neocortical regions with special structure (e.g. primary sensory areas). This classification much abbreviates the analysis (§ 4.7) of the rest of the hippocampal formation.

4.6.1. *Memorizing cortex*

M 1. Large pyramidal cells without climbing fibres, with baskets. These cells usually form simple representations (i.e. can support a collateral effect): they have Brindley afferent synapses, and probably some dendritic independence. It is useful to refer to them as memorizing cells.

M 2. Star cells, and small pyramidal cells without climbing fibres, with baskets, are like **M 1.** They may be used with baskets in a simple memory, where subevents not wholly included in a learnt event are used to address that event, and are also included in the term 'memorizing cell'.

M 3. Star cells or small pyramids, without baskets, without climbing fibres, with small dendrites and ascending axons, are codon cells, used only at the first stage of a simple memory. Perhaps with modifiable synapses (Brindley), their principal function is to reduce α.

M 4. Short-axon cells, without afferent baskets, without climbing fibres, with small dendrites, driven mainly by **M 1** or **M 2** cell collaterals, and with ascending axons. These cells are inhibitory. They control **M 1**, **M 2** or **M 3** cell dendritic thresholds for synaptic modification, and the level of subtractive inhibition during recall.

M 5. Short-axon cells like **M 4** only with local axons and dendrites. They synapse with **M 1**, **M 2** or **M 3** cells, and are inhibitory.

M 6. Basket cells, driven by the same afferents as drive **M 1, 2 or 3** cells, and sending inhibitory synapses to the somas of these cells. Basket cells may also receive synapses from **M 1** to **3** cell axon collaterals, since this would be one way of raising f during recall.

M 7. Fusiform cells lying deep in the cortex, with a liberal dendritic expansion and local axonal arborization, typically to **M 3** or **M 1** and **2** cell dendrites. They are inhibitory threshold controlling cells, like **M 4**, which operate by negative feedback to the cells whose thresholds they control, and by direct sampling of afferents (cf. cerebellar Golgi cells).

4.6.2. *Classifying cortex* (Marr 1970)

C 1. Pyramidal cells with afferent climbing fibres and basket synapses, are cells representing classificatory units.

C 2. Star cells, or granule cells, without baskets, without climbing fibres, with small dendrites and often an ascending axon, are codon cells. They are driven mainly by afferents to that region of cortex, and some may have modifiable afferent synapses.

C 3. Cells whose axons become climbing fibres.

C 4. Short-axon cells other than **C 2**, with local axonal and dendritic ramification: they are inhibitory.

C 5. Basket cells, similar to **M 6.**

C 6. Fusiform cells with single ascending and descending dendritic shaft, usually lying deep in the cortex, and possessing an axon that goes to white matter without emitting any collaterals. These cells are probably cortical indicator cells of some kind, and some may project to archicortex.

4.7. *The histology of various hippocampal areas*

The letters (e.g. **M 3**) accompanying the following descriptions of the histology of allocortical regions refer to the cell classifications of §4.6. No detailed justifications of these diagnoses are given, since the arguments used for such justifications have all appeared in §4.

4.7.1. *The area entorhinalis* (a.e.)

The a.e. was studied by Cajal (1911) and by Lorente de No (1933), who reviewed and revised Cajal's work. The following summarizes the account given by Lorente de No (1933), which combines his and Cajal's work. Roman numerals indicate cortical layers, taken after Lorente de No.

I. Plexiform layer, with the usual short-axon cells (**M 5**). The axons here are mainly ascending axons from deeper layers (e.g. from layer V), and association fibres from other fields arriving through the plexiform layer.

II. Layer of star cells (**M 2**): their axons are thick and go to the white matter after giving off many collaterals. There are also various short-axon cells, some of which may synapse with the star cell somas (**M 5, M 4,** possibly **M 6**).

III. Layer of superficial pyramids (**M 2**). These cells have many dendrites in I, no branching in II, and a dense basilar dendritic field. The cingulum afferents to a.e. seem to end among these basilar dendrites (White 1959). The axon sends collaterals mainly to I and III (some to II and V) and goes to the white matter. Various short axon and miscellaneous other types of cell (**M 4** to **7**) are also found (III includes Cajal's (1911) layer 4°).

IV. Layer of deep pyramids, with thin unbranched dendritic shaft and immense basilar dendritic plexus (**M 2**). In this layer it is indigenous dendrites, rather than foreign axons, which

arborize and ramify. Their axons project to the white matter giving off many collaterals to I, II, III and V. The ascending collaterals rise vertically. Horizontal cells are also found here, probably including basket cells, and various cells with ascending axon (**M 4** to **7**). No collaterals of any extrinsic afferents terminate in this layer.

V. Small pyramidal cells with recurrent axons (**M 3**). Their axons send collaterals to I, II, III and V but not to IV. In IV, however, the dendrites ramify profusely, and the ascending axons synapse with them (probably) forming their main source of afferents. Globular cells with long dendrites inhabit layers V and VI, their axons arborising densely in layer V or VI (**M 7**). Spindle cells with short axons and local dendrites (**M 4, 5**) are also found. According to Cragg (1965), it is the fibres from ventral temporal neocortex which terminate here, in the cat.

VI. Layer of polymorph cells: there are many types, none particularly surprising; globular, polygonal, and those left over from V. They have various combinations of axonal and dendritic distributions (**M 3** to **7**).

4.7.2. *The presubiculum*

Cajal (1911) is the only author who has written about the presubicular histology, though Lorente de No (1934) was clearly familar with this area from his own observations (p. 137). It appears that on histological grounds, the hippocampal formation should be divided into three large regions, the Regio Entorhinalis, Regio Presubicularis and Regio Ammonica (Lorente de No 1934, p. 137). The Regio Entorhinalis and the Regio Presubicularis, in spite of many changes—particularly the introduction of star cells to layer II of a.e.—have the same fundamental plan. The Regio Ammonica starts with the introduction of the Ammonic pyramids in layer II of the prosubiculum, and continues into CA and FD. Thus the subiculum may be regarded as transitional cortex (Lorente de No's Subiculum *b*). (Cajal took what Lorente de No calls presubicular cortex (Sub. *a*) for his description of the human subiculum.)

The division of the hippocampal formation into three large areas, as suggested by Lorente de No on histological grounds, will be adopted here. The argument will essentially be that the Regio Entorhinalis and the Regio Presubicularis prepare information from many different sources for its simple representation in the CA and FD. It seems probable that each collection of cells in the Regio Entorhinalis and the Regio Presubicularis should be treated as preparing information from a separate source: the different shapes of the cells reflect the particular statistical quirks of the different kinds of information. The layer $\mathscr{P}_2$ of § 3.1 is a rough model for all of them.

The lack of detailed information about the Regio Presubicularis prevents its detailed discussion. The presubiculum of Cajal (1911) is presented as a typical example of presubicular cortex.

I. Plexiform layer, extremely wide, and containing many afferents to CA and FD. Its outer zone is composed almost entirely of such fibres, but the inner part contains the terminal bushes of ascending dendrites from layers described below, and so is a true plexiform layer. This region presumably contains the usual short-axon cells (**M 5**), but they seem to be difficult to stain with the Golgi method (Lorente de No 1934).

II. Layer of small pyramids and fusiform cells (**M 2, 3, M 7**?). The axons of many of these cells descend to the white matter, some ending locally. The dendrites of all seem to be confined to layers I and II.

III. Deep plexiform layer. (Lorente de No might have combined II and III as he did in a.e.) This layer is thick, with relatively few cells; small and medium pyramids (**M 2, 3**?) and various other cells (**M 4, 5, 7, 6**?). It contains an extremely dense plexus, and apparently, the layers I

to III receive here the terminal ramification of the massive pathway to the presubiculum carried by the cingulum.

IV. Large and medium pyramidal cells (**M 1, 2?**): the smaller pyramids are probably on average lower in the cortex, and their basilar dendrites generate a dense horizontal plexus. The large ones seem to have a more irregular dendritic arrangement (though information is very sparse, and these statements are inferences from Lorente de No's (1934) incidental remarks). All pyramidal cell axons go to the white matter. The large pyramids of this layer become layer III of the prosubiculum, and seem to be associated with the existence of Martinotti type cells (**M 3**) beneath them.

V. Fusiform and triangular cells, similar to those found in other cortical areas (**M 4, 5, 7**), and cells with ascending axon (**M 3**). No details are available.

4.7.3. *The subiculum* (Prosubiculum + Sub. *b* of Lorente de No)

It is convenient in this section to use the terminology of Lorente de No (1934, p. 134).

The subiculum lies next to CA 1, into which it gently merges. Only a very small region (Lorente de No's Sub. *b*) can be said to have a distinctive structure in that the presubicular pyramids have disappeared, but the prosubicular pyramids have not yet appeared. The huge terminal ramification of the cingulum is strictly confined to the presubiculum, and does not spill over into the prosubiculum (White 1959; Raisman *et al.* 1965; Cragg 1965).

I. An extremely wide plexiform layer, containing the perforant tract from a.e. to CA and FD. The lower zone is a true plexiform layer, and contains horizontally running collaterals of some of the fibres running overhead. There are the usual short-axon cells (**M 5**).

II. Modified ammonic pyramids (**M 1**). The apical dendrites lack S. Lac. and S. Rad., which ceases abruptly at the edge of CA 1. The basal dendrites are horizontal, and none descend to III. There are also many short axon and basket cells (**M 4, 5, 6**).

III. Prosubicular pyramids: the upper cells have no side branches in III to their dendritic shafts, but the lower ones do. None have any in II; all have them in I. Thus the cells of II avoid the plexus in III, and the cells of III avoid the plexus in II. These cells are probably **M 1**. Again, there are various short axon and basket cells (**M 4** to **6**).

Many pyramidal cells in the prosubiculum send axon collaterals to CA 1 and CA 2. Most axons enter the alveus of the CA, and thence enter the fimbria.

IV has two strata: (*a*) of globular cells, of which there are various kinds. Those whose axons pass to the white matter are probably **M 1**, and those with ascending axon are probably **M 3**; and (*b*) of Martinotti (**M 3**) type cells with local dendrite and ascending axon. These seem to be associated with the prosubicular pyramids, and to die out with them, which suggests that their axons do not rise above III. It may be these axons which cells of II are anxious to avoid.

Layer IV, especially IV b, becomes very thin towards the CA. III becomes very wide, and the cells seem to turn into Ammonic pyramids as IV b disappears (Lorente de No 1934, p. 129, figure 11). The prosubiculum thus merges into and becomes the CA, which has already been described.

5. Neurophysiological predictions of the theory

5.0. *Introduction*

In this section are summarized the most important predictions which follow from the notion that simple memory provides a model for the archipallium in general, and for the hippocampus in particular. They are presented in two parts; the first summarizes the general model for archicortex, and the second deals with the detailed predictions for the hippocampus.

The statements are made with varying degrees of firmness, which are indicated by the number of stars accompanying each (after Marr 1970, §7). Three stars indicates a prediction whose disproof would show simple memory theory to be an inappropriate model for archicortex; a no-star prediction is a strong hint and nothing more: one and two star statements lie between these extremes.

5.1. *The general model for archicortex*

Whereas neocortex is capable both of classifying and of memorizing inputs, archicortex is capable only of memorizing them***. The variety of the functions performed by archicortex is achieved in part by the application of its basic memorizing ability to widely different kinds of information. Two examples of the uses to which archicortex may be put are free simple memory (in which the memory projects to its own input cells), and directed simple memory (in which it does not).

The central feature of archicortex is a collection of so-called memorizing cells, identified as that class of cell which is most numerous and whose axons project elsewhere. Such cells will have at least two kinds of afferent synapses***: excitatory afferent synapses with Brindley modification conditions***; and unmodifiable inhibitory afferent synapses***. The dendrites of memorizing cells are often independent**, modification conditions being decided locally**.

The inhibition applied to memorizing cells performs at least two principal functions: one is to control the synaptic modification conditions in the memorizing cell dendrites during the learning of events**; and the other is to control the cells' thresholds during the recall of previously learned events***. Cells for the first function apply inhibition to the dendrites of the memorizing cells**, and are driven either by memorizing cell axon collaterals, or by afferent collaterals, or both (by analogy with the cerebellar Golgi cells). They act so as to maintain the number of memorizing cells involved in learning each new event at a roughly constant level**.

Cells for the second function are of two types**; basket cells, performing a division**, and stellate cells, synapsing with the dendrites, performing a subtraction**. The stellate cells act to remove from the output signal some of the excitation due to the unmodifiable component of the Brindley synapses*. The basket cells and stellate cells are driven by the main afferent system to the memorizing cells (through unmodifiable excitatory synapses)**, and perhaps also by memorizing cell collaterals*. It is appropriate in certain circumstances to raise the division threshold of the memorizing cell during recall of a learnt event**. There are various circuits capable of achieving this.

Archicortex may contain codon cells, perhaps with modifiable afferent synapses. If so, and if the synapses are modifiable, then they are Brindley synapses**, and are accompanied by the same kinds of inhibitory housekeeping cells as are memorizing cells**. They are often small and numerous**, and are necessary when the activity (α) of the input fibres is too high for the learning capacity required of the memorizing cells***.

It is the lack of climbing fibres which deprives archicortex of the clustering ability underlying

the classification process in neocortex**. Archicortex is therefore bad at the kind of classification of which neocortex is probably capable***.

This outline of the processes carried out in archicortex gives rise to a rough classification of archicortical cell types. These have been labelled **M 1** to **M 7**, and are not set out here since they have been summarized in the appropriate way in § 4.6.1. For the purposes of this section, they may be regarded as owning two stars, except where overridden by the statements made above.

5.2. *The hippocampal cortex*

Star ratings in this section test the proposition that the various divisions of the hippocampal formation form components of a simple memory.

The pyramidal cells of CA 1 to 3 and the granule cells of the FD are memorizing cells, in the sense of § 4.6.1***. Their main afferents therefore terminate by means of Brindley modifiable synapses***. All other cells there are probably inhibitory**, and certainly many are***. These cells are concerned with the formation of simple representations, in the sense of § 3**. That is, the activities of these populations are low** (near 0.001) and there is an extensive collateral system** which uses Hebb (or Brindley) modifiable synapses**. The collaterals aid the completion of simple representations during recall**. The performance of regions of the CA (e.g. say CA 2) is qualitatively similar to that of the layer $\mathscr{P}_3$ in the explicit model of § 3.1**.

The star cells of the entorhinal area are also memorizing cells***, and are qualitatively analogous to the layer $\mathscr{P}_2$ of the model of § 3.1**. Various predictions follow from these remarks, in particular that they possess Brindley modifiable afferent synapses***. Many other cells in various archicortical areas have been discussed, and the predictions concerning them follow the general lines of § 5.1. In the following lists, the various cells are classified according to the terminology of § 4.6.1; the firmness of the classification is indicated; and the references specify the relevant pieces of text.

5.2.1. *Cornu ammonis*: *CA* 1 *to* 3

cell type	described (§)	stratum	class	reference(§)	stars
pyramid	4.1.1 (*a*)	pyr.	**M 1 or 2**	**4.2**	***
pyr. basket	4.1.1 (*b*)	pyr.	**M 6**	**4.3.3**	***
asc. axon	4.1.1 (*c*)	pyr.	**M 4, 5**	**4.3.2**	**
horizontal	4.1.1 (*d*)	oriens	**M 4**	**4.3.2**	**
polygonal	4.1.1 (*e*)	oriens	**M 4, 5**	**4.3.2**	**
basket	4.1.1 (*f*)	oriens	**M 6**	**4.3.3**	***
basket	4.1.1 (*g*)	oriens	**M 6**	**4.3.3**	***
horizontal	4.1.1 (*h*)	oriens	**M 4**	**4.3.4**	**
horizontal	4.1.1 (*i*)	oriens	**M 5**	**4.3.2**	**
various	4.1.1 (*j*)	rad. & lac.	**M 5**	**4.3.2**	***
asc. axon	4.1.1 (*k*)	rad. & lac.	**M 4**	**4.3.2**	***
horizontal	4.1.1 (*l*)	rad. & lac.	**M 5**	**4.3.4**	***
short axon	4.1.1 (*m*)	molec.	**M 5**	**4.3.2**	***
horizontal	4.1.1 (*n*)	molec.	**M 5**	**4.3.2**	***

One kind of cell can fall into two classes if it possesses two kinds of dendritic or axonal distribution.

There may be an afferent system capable of changing the ratio of somatic to dendritic inhibition at the CA pyramids. This would increase the amount of basket inhibition during recovery of a simple representation. No-star estimates of the values of the relevant parameters for CA appear in § 4.2.4.

5.2.2. *The fascia dentata*

cell	described (§)	layer	class	reference (§)	stars
granule	4.1.2	granular	M 1 or 2	4.4.1	***
displaced gran.	4.1.2 (*a*)	molec.	M 1 or 2	4.4.1	**
short axon	4.1.2 (*b*)	molec.	M 4, 5	4.4.1	***
pyr. basket	4.1.2 (*c*)	polymorph	M 6	4.4.1	***
asc. axon	4.1.2 (*d*)	polymorph	M 4	4.4.1	**
desc. axon	4.1.2 (*e*)	polymorph	} probably CA 4.		
short axon	4.1.2 (*f*)	polymorph			
star, etc.	4.1.2 (*g*)	polymorph			

α_{FD} is probably rather low (near 0.001)*.

5.2.3. *CA* 3, *CA* 4 *and the mossy fibres*

The pyramids of CA 4 are 'collector' cells for the output of FD granule cell activity*, (§§4.4.0, 4.4.3). They may have Brindley modifiable afferent synapses from FD granule cell axons*, being the short-axon cells of CA 4 the necessary class M 4 and M 5 cells*. The mossy fibre synapses in CA 3 may be Hebb or Brindley synapses*. The large size of the mossy fibre synapses suggests that α_{FD} is very low*—certainly lower than α for the other hippocampal afferents** (§4.4.3).

5.2.4. *Hippocampal collateral systems*

All short hippocampal pyramidal cell collaterals to other hippocampal pyramids end in Hebb or Brindley modifiable synapses**. Those collaterals which are reciprocated can take part in the collateral effect**. Those which do not are concerned with associating simple representations formed in different regions of the hippocampus (§4.5.1), these being completed by local reciprocating collaterals*. Examples of the second sort are the mossy fibres**, and the Schaffer collaterals**. Examples of the first kind are local collaterals**, and perhaps commissural connexions (§4.5.2). There should be $\gtrsim$ 10000 collateral synapses at each CA pyramidal cell**. Local collaterals joining hippocampal pyramids tend to make synapses in S. Rad.* (§4.5). There may be a collateral effect in FD (§4.5.3).

5.2.5. *Area entorhinalis*

cell	described (§)	layer	class	references (§)	stars
short axon	4.7.1 I	I	M 5	4.7.1	***
star	4.7.1 II	II	M 2	4.7.1	***
various	4.7.1 II	II	M 4, 5, 6?	4.7.1	**
pyramid	4.7.1 III	III	M 2	4.7.1	***
various	4.7.1 III	III	M 4–7	4.7.1	**
pyramid	4.7.1 IV	IV	M 2	4.7.1	***
various	4.7.1 IV	IV	M 4–7	4.7.1	**
pyramid	4.7.1 V	V	M 3	4.7.1	***
globular	4.7.1 V	V	M 7	4.7.1	**
spindle	4.7.1 V	V	M 4, 5	4.7.1	**
polymorph	4.7.1 VI	VI	M 3–7	4.7.1	*

5.2.6. *Presubiculum*

cell	described (§)	layer	class	references (§)	stars
short axon	few seen	I	**M 5**	4.7.2	***
pyramids	4.7.2 II	II	**M 2 or 3**	4.7.2	***
fusiform	4.7.2 II	II	**M 7**	4.7.2	none
various	4.7.2 III	III	**M 2–7**	4.7.2	(little information)
pyramids	4.7.2 IV	IV	**M 1?, M 2**	4.7.2	**
fusiform, triangular	4.7.2 V	V	**M 4, 5, 7**	4.7.2	*
asc. axon	4.7.2 V	V	**M 3**	4.7.2	**

This region has been studied even less than the others.

5.2.7. *Prosubiculum* (of Lorente de No)

cell	described (§)	layer	class	reference (§)	stars
short axon	4.7.3 I	I	**M 5**	4.7.3	***
pyramid	4.7.3 II	II	**M 1 or 2**	4.7.3	***
short axon	4.7.3 II	II	**M 4, 5**	4.7.3	**
basket	4.7.3 II	II	**M 6**	4.7.3	***
pyramid	4.7.3 III	III	**M 1 or 2**	4.7.3	***
pyramid	4.7.3 IV	IVa	**M 1 or 2**	4.7.3	***
Martinotti	4.7.3 IV	IVb	**M 3**	4.7.3	***
short axon	4.7.3	III, IV	**M 4–6**	4.7.3	**

The prosubicular pyramids are probably **M 1 * *** since they send collaterals to CA and axons to the fimbria.

I wish to thank Professor G. S. Brindley, F.R.S. for his helpful criticisms, and Mr S. J. W. Blomfield for many discussions. The following kindly gave me permission to reproduce figures from other papers: Dr T. W. Blackstad and the Wistar Press for figure 2; Professor R. Lorente de No and Akademic-Verlag GmbH for figures 8 to 11; and C.S.I.C. Madrid for figures 12 to 16. The work was supported by Trinity College, Cambridge.

Note added in proof, 15 *April* 1971

Lømo (1971) has published evidence for the facilitation of the perforant path—Dentate granule cell synapses in the rabbit. His findings are necessary but not sufficient to prove this theory's prediction (§5.2.2) that these synapses are facilitated by simultaneous pre- and post-synaptic depolarization.

References

Andersen, P. O. 1966 Correlation of structural design with function in the archicortex. In *Brain and conscious experience* (ed. J. C. Eccles), pp. 59–84. Berlin: Springer-Verlag.

Andersen, P. O., Eccles, J. C. & Løyning, Y. 1963 Recurrent inhibition in the hippocampus with identification of the inhibitory cell and its synapses. *Nature, Lond.* **198**, 540–542.

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.

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

Cajal, S. R. 1911 *Histologie du Système Nerveux*, Tome II. Madrid: C.S.I.C.

Cragg, B. G. 1965 Afferent connections of the allocortex. *J. Anat.* **99**, 339–357.

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

Eccles, J. C., Llinás, R. & Sasaki, K. 1966 Parallel fibre stimulation and the responses induced thereby in the Purkinje cells of the cerebellum. *Expl Brain Res.* **1**, 17–39.

Hamlyn, L. H. 1962 The fine structure of the mossy fibre endings in the hippocampus of the rabbit. *J. Anat.* **96**, 112–120.

Lømo, T. 1971 Potentiation of monosynaptic EPSP's in the perforant path—dentate granule cell synapse. *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.* **202**, 437–470.

Marr, D. 1970 A theory for cerebral neocortex. *Proc. Roy. Soc. Lond.* B **176**, 161–234.

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

Raisman, G., Cowan, W. M. & Powell, T. P. S. 1966 An experimental analysis of the efferent projection of the hippocampus. *Brain* **89**, 83–108.

Schaffer, K. 1892 Beitrag zur Histologie der Ammonshornformation. *Arch. mikrosk. Anat.* **39**, 611–632.

Shariff, G. A. 1953 Cell counts in the primate cerebral cortex. *J. comp. Neurol.* **98**, 381–400.

Spencer, W. A. & Kandell, E. R. 1961 Electrophysiology of hippocampal neurons. IV. Fast prepotentials. *J. Neurophysiol.* **24**, 274–285.

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

David Willshaw

Commentary on

Simple Memory: A Theory of the Archicortex

This is the third, and last, of David Marr's series of three theoretical papers on the neurobiology of learning and memory (Marr 1969, 1970, 1971). In this paper, he proposes a theory for the functioning of the mammalian hippocampus — one of the most important but least understood parts of the brain.

The hippocampus is one of the phylogenetically older parts of the brain (hence:archicortex). It is found in mammals, and a related structure exists in birds. The mammalian hippocampus has a simple and regular structure, and specific circuits have been identified within it. It has afferent and efferent pathways to many parts of the neocortex, and these interconnections are fairly well characterized.

It has proved difficult to assign positively any definite function, or functions, to the hippocampus. Nonetheless, various proposals have been made. At the time Marr wrote this paper, the startling results from such patients as HM, who became amnesic after undergoing bilateral hippocampectomy for the relief of epilepsy, suggested a role for the hippocampus in memory (Scoville and Milner, 1957). More recently, the idea has been developed that a "cognitive map" is built in the hippocampus (O'Keefe and Nadel, 1978). This is based on the finding that there are "place units" in the rat hippocampus — neurons that fire when the animal is at a specific place in the environment.

Marr had previously proposed (1970) that the neocortex is the site of long-term associative storage of information, the information being stored in a form that retains the essential details and removes the superfluous. In the hippocampus paper, he argues that it would be inefficient to store the raw associations directly, before the salient features had been extracted; furthermore, neocortical interconnectivity is not sufficiently complete to allow any arbitrary association to be stored. Marr proposes that there is a special temporary memory store — the hippocampal formation.

The central question is concerned with the architecture required for this temporary memory, and whether it matches the known structure of the hippocampus. As in his other papers on learning and memory, Marr's method of working is to constrain the problem by a number of assumptions as to the likely values of some of the parameters of the system. These values either were derived intuitively (e.g., the number of items to be stored) or had some biological basis (e.g., the number of synapses on a nerve cell). To these assump- tions are added a number of computational constraints that must hold if the memory is to perform satisfactorily. He concludes that there

must be an intermediate layer of cells between the input and output layers of the memory. In the present day parlance of Connectionism, this would have a natural interpretation as a layer of hidden units. Having derived a complete specification of this three-layer system, he goes on to relate this three-layer model to the known facts of the hippocampus and its connections to the neocortex.

In this paper, Marr's use of a set of constraints to derive the minimal structure for the given problem reaches a most sophisticated level. However, his attempt is not wholly satisfactory, since there is an inconsistency in the argument, which leaves his case for a three-layer model not proven. He therefore relies more heavily on his view of hippocampal circuitry than is stated explicitly. In effect, he views the problem from two different perspectives: (a) that the structure of the memory proposed is necessary on computational grounds and (b) that it has to have this structure because this is the way that the hippocampus was built. This double perspective can be seen in light of his subsequent development of the importance of the computational, the algorithmic and the implementational levels of explanation (Marr, 1982).

Although he does characterize in some detail the individual properties of the cells that are meant to form the layers of his model, only a loose correspondence is made between the subdivisions of the hippocampus (together with the associated neocortical circuitry) and the layers of his model. The most extensive discussion is concerned with the nature of the cells of the output layer of the memory, which are identified with the pyramidal cells of the hippocampus. He does not distinguish between the various elements of the Dentate Gyrus-CA3-CA1 trisynaptic circuit, the existence of which was known at the time (Andersen et al., 1971). This may have been a foresighted omission, given that the notion of the trisynaptic circuit itself is now in the process of change with the discovery of other extrinsic pathways of the hippocampus (Squire et al., 1989). His major contribution is in his discussion, at a cellular and subcellular level, of the properties that the individual elements of his model must have. In particular, he proposes various types of *dual strategies* for setting the thresholds of the cells (which have never been properly investigated since), which are required for efficient storage and retrieval in the biologically realistic cases of incompletely connected networks. The requirement that synapses be modified by simultaneous presynaptic and post-synaptic activity, after the style of Hebb (1949), predates the discovery of hippocampal long-term potentiation (Bliss and Lømo, 1973), although he does add a note in proof about Lømo's earlier paper (1971) that showed synaptic facilitation in the perforant path — dentate gyrus.

In summary, David Marr presents a somewhat abstract interpretation of the hippocampus as a temporary memory store. The strength of his analysis lies not in the translation of his formal model into neurobiological terms, but rather in his discussion of what types of local circuitry are required to perform the various computations that are needed for the memory to function efficiently.

It is unfortunate that this paper is not more widely read or understood. It

required considerable effort to come to terms with the inaccessible style that is characteristic of his earlier writings; but I found that the effort was well worthwhile. Even 20 years after publication, Marr's theory remains the most complete computational model of the hippocampus.

This short commentary is based on a recent review of the computational basis of Marr's theory of archicortex (Willshaw and Buckingham, 1990). We also describe the results of analysis and of computer simulations that were designed to compare the performance of two-layer and three-layer models.

REFERENCES

Andersen P, Bliss TVP, Skrede K (1971): Lamellar organization of hippocampal excitatory pathways. *Exp Brain Res* 13:222-238

Bliss TVP, Lmo T (1973): Long-lasting potentiation of synaptic transmission in the dentate area of the anesthetized rabbit following stimulation of the perforant path. *J Physiol* 232:331-356

Hebb DO (1949): *The Organization of Behavior*. New York: John Wiley and Sons

Lømo T (1971): Potentiation of monosynaptic EPSP's in the perforant path— dentate granule cell synapses. *Exp Brain Res* 12:46-63

Marr D (1969): A theory of cerebellar cortex. *J Physiol* 202:437-470

Marr D (1970): A theory for cerebral cortex. *Phil Trans Roy Soc* B 176:161-234

Marr D (1971): Simple memory: theory for archicortex. *Phil Trans Roy Soc B* 262:23-81

Marr D (1982): *Vision*. New York: Freeman

O'Keefe J, Nadel L (1978): *The Hippocampus as a Cognitive Map*. Oxford: Oxford University Press

Scoville WB, Milner B (1957): Loss of recent memory after bilateral hippocampal lesions. *J Neurol Neurosurg Psychiat* 20:11-12

Squire LR, Shimamura AP, Amaral DG (1989): Memory and the hippocampus. In: *Neural Models of Plasticity*, Byrne J, Berry W, eds. NY: Academic Press

Willshaw DJ, Buckingham JT (1990): An assessment of Marr's theory of the hippocampus as a temporary memory store. *Phil Trans Roy Soc B* (in press)

Professor
Centre for Cognitive Science
University of Edinburgh
Edinburgh, Scotland UK

Bruce McNaughton

Commentary on

Simple Memory: A Theory of the Archicortex

I regard it as a significant honor to be able to comment here, from a neurobiologist's perspective, on the impact of David Marr's theoretical neural network models on our understanding of the biology of associative memory, in particular in the mammalian hippocampal formation and neocortex. While there is some truth to Willshaw and Buckingham's (1990) suggestion that some of us have cited Marr's papers rather more widely than we have understood them, his three papers (Marr 1969, 1970, 1971) on the cerebellum, neocortex and archicortex (hippocampus) have been guiding lights both to myself and to a number of other experimental neuroscientists. (It is unfortunately also the case that Marr's ideas are sometimes more widely exploited than they are cited.) Marr's approach, in its mathematical rigor, was always difficult, and often obscure to the non-mathematician. This, unfortunately, led to his theories being less widely appreciated (or understood) among neurobiologists than they might otherwise have been; however, the value of Marr's models for neurobiological studies lies not so much in their mathematical sophistication or overall correctness in detail (they are almost certainly wrong), but for the far-reaching explanatory power of their relatively simple individual components. It is the broad conceptual framework provided by these models, rather than their correctness in detail, that will insure Marr his important place in the historical development of our understanding of how biological neural networks actually operate. Looking back to the sparsity of the experimental database from which Marr developed his ideas, it is astounding the extent both to which these insights have been substantiated, and to which they have brought order to a number of otherwise disconnected data on the anatomy, biophysics and information transmission of the mammalian hippocampal formation and its relations with the neocortex. Contrary to Willshaw and Buckingham's (1990) statements, many of Marr's predictions have, in fact, been followed up. In the following I shall attempt to illustrate this with a few of the more salient examples.

Synaptic Modification

Marr was the first theoretician to attempt to make use, in the context of a detailed, neurobiologically constrained model, of Hebb's postulate that synapses should be enhanced under conditions of conjoint pre- and postsynaptic depolarization. At the time that he wrote, the first experiments by Lømo, and subsequently by Bliss, Gardner-Medwin and Lømo, were beginning to reveal that hippocampal synapses exhibited a plasticity of sufficient duration to be considered as a candidate for associative memory; however, it was not until much later that the first evidence was obtained that Hebb's

principle might be implemented in this process (McNaughton, Douglas and Goddard, 1978), and later still before this was fully confirmed and understood mechanistically (Collingridge et al., 1983; Gustafsson et al.,1987; Harris et al., 1984; Kelso et al., 1986; Wigström et al 1986). A substantial body of literature has accumulated that is at least consistent with the idea that this process does, indeed, reflect the experimental activation of mechanisms that normally subserve at least the initial registration of associative memories (see McNaughton and Morris, 1987, for overview). Most of the available data indicate that the characteristics of the main modification process are consistent with what Marr called "Brindley" synapses (which have a non-modifiable excitatory component) rather than binary "Hebb" synapses, although this question is by no means closed.

Pattern Completion

In the archicortex paper, Marr suggests that the completion of stored events from fragmentary or noisy input information should be the primary function of the "simple memory" system he envisioned for the hippocampus. This fundamental idea has proven to be of immense value in the design of neurophysiological and behavioral experiments, and two lines of investigation now suggest the fundamental correctness of this assertion. In the rodent hippocampus, the "events acted upon relate primarily (or at least most obviously) to the animal's representation of space. Individual pyramidal cells are selectively active in limited regions and orientations within the animal's known environment. Although these "place fields" are determined by the animal's orientation with regard to the distal visual landmarks, removal of any subset of these landmarks has little or no effect on the spatial information transmitted by these cells (O'Keefe, 1976; O'Keefe and Conway, 1978). More direct evidence for pattern completion in hippocampal circuits was recently obtained in studies (Mizumori et al., 1989b) in which the discharge rates and spatial selectivities of CA3 pyramidal cells were severely curtailed by temporary inactivation of a modulatory input from the medial septum, which is necessary to maintain the excitability of CA3 cells. Pyramidal cells in CA1, whose major source of modifiable excitatory input is CA3, were almost completely unaffected. Somehow, the highly reduced subsets of spatial representations conveyed from CA3 were sufficient to enable complete spatial representations in CA1.

Inhibitory Control of Global Threshold During Storage and Recall

Perhaps the most insightful and powerful of Marr's ideas was his suggestion that inhibitory interneurons should control both the threshold for synaptic modification during storage, and, by means of a division operation, the output threshold for principal cells during associative recall (pattern completion). The former notion has been verified in a number of studies that have shown that the modification of hippocampal synapses is largely regulated by GABAergic inhibition (Wigström and Gustafsson, 1983; Sharfman & Sarvey, 1983; Larson et al., 1986). The latter idea, although more difficult to verify, has some experimental support. Inhibition mediated by the chloride dependent $GABA_A$ channel is fundamentally a division operation (for elaboration, see McNaughton and Barnes, 1990, and McNaughton and Nadel, 1989). Because the chloride equilibrium potential is almost the same as the resting potential, the effect

of inhibition (relative to rest) is primarily to increase membrane conductance. Because the soma voltage change is roughly the outward excitatory synaptic current (i_m) divided by membrane conductance (g_m), and because resting conductance is small, a division operation is implemented ($DV_m = i_m/g_m$). Secondly, in the studies cited above by Mizumori et al. (1989b) in which CA1 output was preserved in the face of reduced and degraded CA3 input, the activities of basket inhibitory interneurons were reduced in proportion to the reduced CA3 input. This appears consistent with Marr's idea that inhibitory cells should sample the activity in the input fiber population and feed forward a proportional division signal. Also consistent is the fact that, in all hippocampal subfields, most inhibitory cells receive direct excitation from the same modifiable excitatory inputs that project to the principal cells. As suggested by the idea of setting the output threshold globally, these cells need not be numerous, and indeed, in the hippocampus, they constitute only a small population relative to the principal cells. It is also known that the behavioral conditions under which the density of afferent activity from entorhinal cortex to hippocampus is greatest are also the conditions under which hippocampal inhibitory cells are most active. In further support of this idea, the probability of inhibitory cell output is graded with stimulus intensity (i.e., number of active afferents), whereas the principal cells do not normally fire until the intensity is high enough to activate many more afferents than would be coactive in a typical physiological event (Mizumori et al., 1989a).

Another interesting consequence of the threshold setting hypothesis is that, unlike principal cells, which care about exactly *which* afferents are active in an event, the inhibitory cells should care primarily only about *how many* are active (McNaughton and Nadel, 1990). This clearly characterizes the differences in spatial firing characteristics between hippocampal pyramidal and basket cells.

Finally, although Marr did not consider in detail the dynamics of his proposed 'input normalization', there is one logical consequence of this scheme which provides considerable insight into the dynamics of the feed-forward inhibitory networks of the hippocampus. In order for the division operation to be effective, the division signal arriving at the principal cell soma must arrive with or before the excitatory synaptic signal from the current event; however, the inhibitory signal must cross two synapses, whereas the excitatory signal need cross only one. To compensate for this, the inhibitory system appears to have evolved an extremely rapid response mechanism. When hippocampal afferent fibers are electrically activated, inhibitory cells fire well before principal cells (Ashwood et al., 1984; Buzsaki, 1984; Douglas, McNaughton and Goddard, 1983; Mizumori et al., 1989a) so that the inhibitory conductance in the principal cells is already activated before most principle cells reach threshold.

The Necessity for Keeping a Low

Marr proposed that the simple memory system must satisfy the dual constraints of maximizing the event storage capacity, while at the same time preserving enough information from each event to ensure reliability. These constraints essentially dictated the size of the required networks, and the proportion a of cells that could be used in the representation of any given event. Marr proposed that the value of a should lie between 0.01 and 0.001, and be

roughly constant across events. To translate this into actual neuronal firing rates, take as the 'time-step' the apparent time constant of most hippocampal and cortical neurons, which is on the order of 0.01 sec. The corresponding average firing rates then become between 1.0 and 0.1 Hz, values that are quite low by the standards of most cortical neurons. It turns out that these are about the typical mean discharge frequencies for hippocampal principal cells recorded in alert rats during the performance of spatial learning tasks dependent on the integrity of the hippocampus (O'Keefe, 1976; McNaughton et al., 1983). This 'sparse' encoding of events is also manifested in the exquisite spatial selectivity exhibited by hippocampal pyramidal cells. In extended spaces, a typical cell fires intensely only in a highly restricted region of the animal's known accessible environment, a region typically covering on the order of 0.01 to 0.001 of the total area (these values vary somewhat depending on the size of the environment and other factors). It is also of interest that this sparse coding scheme appears to be a unique characteristic of the hippocampus. In both the entorhinal cortical input and the subicular output structures, spatial coding is considerably more highly distributed, and a (mean firing rate) is correspondingly substantially higher (Barnes et al., 1990).

Marr proposed a rule of thumb for the relationship between a and the number of events n to be stored:

$$n\, a_{i-1}\, a_i \leq 1$$

This ensures that when n inputs have been learned, not all of the synapses have been modified. Using Marr's proposed parameters, this translates to between about 60% and about 10% modified synapses at full capacity, depending partly on how much information is to be made available for retrieval. Above these values, information storage would be unreliable, a given subevent recalling either too many active output cells, or the wrong ones (this is quite analogous to the psychological concept of interference). The prediction of these considerations is that simple memory will fail if the above constraint on the number of modified synapses is exceed. This is exactly the behavioral consequence of artificially increasing the proportion of modified synapses in the hippocampus by high-frequency stimulation of the main input pathways bilaterally. Such stimulation induces a long-term enhancement (LTE/LTP) of a significant proportion of perforant path synapses. This enhancement persists for several weeks, during which time there is both a disruption of recently stored spatial memories and an inability to store new ones (McNaughton et al., 1986; Castro et al., 1990). It is also entirely consistent with Marr's notion of the hippocampus as a temporary memory system that electrically induced synaptic enhancement decays over time, at least at these synapses.

The Collateral Effect

Marr suggested that pattern completion occurred in the pyramidal layers via a "collateral effect". The fundamental idea was that modifiable excitatory collateral synapses would assist recall over several cycles of recurrent excitation. After input of an appropriate subevent, additional cells belonging to the original stored pattern would be activated on succeeding cycles. The "collateral effect" mechanism has now come to be known as "recurrent autoassocia-

tion" (Kohonen, 1972, 1978) and, in one form or another, figures importantly in a number of connectionist style models. Although the implementation of a collateral effect in the hippocampus has yet to be verified experimentally, CA3 has an abundant system of modifiable recurrent collaterals which could perform this function. Also, Marr made two predictions about the dynamics of the collateral effect which seem to be approximately supported by modern data. First, he supposed that about three cycles of the collateral effect should be sufficient to complete the representation. When the hippocampus is actively processing inputs, there is an oscillating cycle of excitation and inhibition known as the theta rhythm, whose mean period is about 140 msec, and to which all hippocampal cell output is phase locked. If one assumes that the completion effect must be going on during the quarter cycle when excitation is increasing, this allows about 35 msec. In the CA3 recurrent system, the combination of conduction delay and synaptic delay amounts to about 6 to 8 msec. This would thus be sufficient for about four to six cycles; only slightly more than Marr predicted. The second prediction was that there should be some external mechanism which gradually increases inhibition during the collateral effect, to make sure only the correct cells were included in the output. The medial septal nucleus, which paces the theta rhythm, has a strong modulatory effect on inhibitory interneurons. As predicted, the activity of the inhibitory cells does increase during the rising excitatory phase of the theta rhythm.

Orthogonalization of Similar Input Vectors

One of the most powerful of Marr's concepts was the idea that memory capacity could be maximized if representations that were rather similar at the input could be recoded by a separate group of cells in such a way as to minimize the overlap in the output. In his cerebellum paper, Marr assigned this function to the cerebellar granule cells, which he called "codon" cells. In the cerebellar paper, the basic idea was to project the input vector onto a higher dimensional space (there are about 40 billion granule cells in the human cerebellum) and then use this orthogonalized representation as input to the memory cells. In the models for neocortex and archicortex, it was considered to be more economical if codons were not hard-wired, but could be created on demand through the use of modifiable synapses. In this way only those codons (subevents) which actually occurred in the experience of the animal would be required. It turns out that the initial projection from the entorhinal cortex into fascia dentata does involve a projection into a higher dimensional space. There are about 105 entorhinal projection cells, and about 106 granule cells in the fascia dentata. This projection terminates in modifiable synapses (probably of the Brindley variety). Moreover, single neuron recording studies of physiologically identified granule cells indicate that a in the granule cell population is among the lowest of any hippocampal subfield (Mizumori et al., 1989a). Thus, although the question requires more systematic investigation, Leonard (1990) has obtained preliminary evidence for pattern separation in the hippocampal output cells in CA1.

Readout from Simple Memory During Sleep

One of the boldest of Marr's predictions was that readout from simple memory should occur during sleep. This idea was first developed in the neo-

cortex paper. Marr argued that "information from which a new classificatory unit is to be formed will often come from a simple associative store," (i.e. hippocampus) "not from the environment ... the most natural way of selecting a location for a new classificatory unit was to allow one to form wherever enough of the relevant fibers converge. This requires that potential codon cells over the whole cerebral cortex should simultaneously allow their afferent synapses to become modifiable. Hence, at such times, ordinary sensory information must be rigorously excluded. The only time when this exclusion condition is satisfied is during certain phases of sleep."

It is unclear whether Marr was aware that at the time this was written, there was a growing psychological literature on the possible role of sleep, particularly REM sleep (Leconte and Bloch, 1971), in memory consolidation (for reviews see Fishbein and Gutwein, 1977; Horne and McGrath, 1984; Smith, 1985). Certainly the basic idea seems to have fallen out from the premises of the model. Recently, some very exciting neurophysiological studies have produced strong support for the plausibility of Marr's idea that information is transferred from temporary (hippocampal) to permanent (cortical) memory during sleep. Pavlides and Winson (1989) studied the effects of selective spatial experience on the subsequent activity of hippocampal "place" cells during sleep. They recorded from pairs of place cells with nonoverlapping place fields. During the waking episode, they exposed the animal to the field of one member of the pair but not to the field of the other. They then removed the animal to a neutral location and allowed it to fall asleep. During the sleep episode, there was a large increase in the output activity of the cells that had been exposed to their place fields, in particular, the occurrence of high-frequency bursts increased, and the interspike intervals during bursts decreased. This are exactly the sort of activity that would be most likely to lead to synaptic modification in target cells. The effect was present in all phases of sleep, but was greatest in REM sleep. This phenomenon thus seems to fit precisely the requirements suggested by Marr's sleep hypothesis.

Closing the Loop

In the foregoing, I have tried to illustrate the astounding prescience of Marr's neurobiological models, and the deep influence his basic ideas either have had or should have on the interpretation of experiments directed towards understanding the different roles of the hippocampus and neocortex in associative memory. Fortunately for the field of computational vision, but unfortunately for the neurobiology of memory, Marr turned his attention away from these problems before completing his theory with a model for the input-output relations between hippocampus and neocortex. He clearly must have thought deeply about this issue, because a forthcoming paper on it was promised but apparently never completed. Many neurophysiologists and neuroanatomists agree that this issue represents the single most important area of almost complete ignorance in the field at present, and Marr's keen insight could very profitably have been applied to this problem. It is amusing to speculate whether, given the rather dramatic increase in our knowledge about the organization of cortical and archicortical memory systems over the past decade, Marr might have turned his attention back once again to these fundamental issues.

REFERENCES

Ashwood TJ, Lancaster B, Wheal HV (1984): In vivo and in vitro studies on putative interneurons in the rat hippocampus: Possible mediators of feed-forward inhibition. *Brain Res* 293:279-291

Barnes CA, McNaughton BL, Mizumori SJY, Leonard BW, Lin L-H (1990): Comparison of spatial and temporal characteristics of neuronal activity in sequential stages of hippocampal processing. *Prog Brain Res* 83:287-300

Bliss TVP, Gardner-Medwin AR (1973): Long-lasting potentiation of synaptic transmission in the dentate area of the unanaesthetised rabbit following stimulation of the perforant path. *J Physiol* 232:357-374

Bliss TVP, Lmo T (1973): Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetised rabbit following stimulation of perforant path. *J Physiol* 232:331-356

Buszáki G (1984): Feed-forward inhibition in the hippocampal formation. *Prog Neurobiol* 22:131-153

Castro CA, Silbert LH, McNaughton BL, Barnes CA (1989): Recovery of spatial learning deficits after decay of electrically induced synaptic enhancement in the hippocampus. *Nature* 342:545-548

Collingridge GL, Kehl SJ, McLennan H (1983): Excitatory amino acids in synaptic transmission in the Schaffer collateral-commissural pathway of the rat hippocampus. *J Physiol (Lond)* 334:33-46

Douglas RM, McNaughton BL, Goddard GV (1983): Commissural inhibition and facilitation of granule cell discharger in fascia dentata. *J Comp Neurol* 219:285-294

Fishbein W, Gutwein BM (1977): Paradoxical sleep and memory storage processes. *Behav Biol* 19:425-464

Gustafsson BH, Wigström WC, Abraham WC, Huang Y-Y (1987): Long-term potentiation in the hippocampus using depolarizing current pulses as the conditioning stimulus to single volley synaptic potentials. *J Neurosci* 7:774- 780

Harris EW, Ganong AH Cotman CW (1984): Long-term potentiation in the hippocampus involves activation of N-methyl-D-aspartate receptors. *Brain Res* 323:132-137

Hebb DO (1949): *The Organization of Behavior.* New York: John Wiley and Sons

Horne JA and McGrath MJ (1984): The consolidation hypothesis for REM sleep function : Stress and other confounding factors—a review. *Biol Psychol* 18: 165-184

Kelso SR, Ganong AH, Brown TH (1986): Hebbian synapses in hippocampus. *Proc Natl Acad Sci USA* 83:5326-5330

Kohonen T (1972): Correlation matrix memories. *IEEE Transactions on Computers*, C-21, Verlag

Kohonen T (1978): *Associative Memory: A system theoretic approach.* New York: Springer-Verlag

Larson J, Wong D, Lynch G (1986): Patterned stimulation at the theta frequency is optimal for the induction of hippocampal long-term potentiation. *Brain Res* 368:347-350

Leconte P, Bloch V (1971): Déficit de in rétention d'un conditionnement après privation de sommeil paradoxal chez le rat. *C R Acad Sci (D) (Paris)* 273:86-88

Leonard B (1990): The contribution of velocity, spatial experience, and proximal visual complexity to the location-and direction-specific discharge of hippocampal complex-spike cells in the rat. Unpublished doctoral dissertation, University of Colorado, Boulder

Lømo T (1966): Frequency potentiation of excitatory synaptic activity in the dentate area of the hippocampal formation. *Acta Physiol Scand (Suppl)* 68:128

Marr D (1969): A theory of cerebellar cortex. *J Physiol* 202:437-470

Marr D (1970): A theory for cerebral cortex. *Phil Trans Roy Soc B* 176:161-234

Marr D (1971): Simple memory: theory for archicortex. *Phil Trans Roy Soc B* 262:23-81

McNaughton BL, Barnes CA (1990): From cooperative synaptic enhancement to associative memory: bridging the abyss. *Sem Neurosci* (in press)

McNaughton BL, Barnes CA, O'Keefe J (1983): The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats. *Exp Brain Res* 52:41-49

McNaughton BL, Barnes CA, Rao G, Baldwin J, Rasmussen M (1986): Long-term enhancement of hippocampal synaptic transmission and the acquisition of spatial information. *J Neurosci* 6:563-571

McNaughton BL, Douglas RM, Goddard GV (1978): Synaptic enhancement in fascia dentata: co-operativity among coactive afferents. *Brain Res* 157:277-293

McNaughton BL, Morris RGM (1987): Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends Neurosci* 10:408-415

McNaughton BL, Nadel L (1990): Hebb-Marr networks and the neurobiological representation of action in space. In: *Neuroscience and Connectionist Theory*, Gluck MA, Rumelhart DE, eds. New Jersey: Lawrence Erlbaum Associates

Mizumori SJY, McNaughton BL, Barnes CA (1989a): A comparison of supramammillary and medial septal influences on hippocampal field potentials and single-unit activity. *J Neurophysiol* 61:15-31

Mizumori SJY, McNaughton BL, Barnes CA, Fox KB (1989b): Preserved spatial coding in hippocampal CA1 pyramidal cells during reversible suppression of CA3 output: evidence for pattern completion in hippocampus. *J Neurosci* 3915-3928

O'Keefe J (1976): Place units in the hippocampus of the freely moving rat. *Exp Neurol* 51:78-109

O'Keefe J, Conway DH (1978): Hippocampal place units in the freely moving rat: why they fire where they fire. *Exp Brain Res* 31:573-590

Pavlides C, Winson J (1989): Influences of hippocampal place cell firing in the wake state on the activity of these cells during subsequent sleep episodes. *J Neurosci* 9:2907-2918

Smith C (1983) Sleep states and learning: a review of the animal literature. *Biobehav Rev* 9:157-168

Scharfman HE, Sarvey JM (1983): Inhibition of post-synaptic firing in the hippocampus during repetitive stimulation blocks long-term potentiation. *Soc Neurosci Abstr* 9:677

Wigström H, Gustafsson B (1983): Large long-lasting potentiation in the dentate gyrus *in vitro* during blockade of inhibition. *Brain Res* 275:153-158

Wigström H, Gustafsson B (1984): A possible correlate of the postsynaptic condition for long-lasting potentiation in the guinea pig hippocampus in vitro. *Neurosci Lett* 44:327-332

Wigström H, Gustafsson B, Huang Y-Y, Abraham WC (1986): Hippocampal long-term potentiation is induced by pairing single afferent volleys with intracellularly injected depolarizing current pulses. *Acta Physiol Scand* 126:317-319

Willshaw DJ, Buckingham JT (1990): An assessment of Marr's theory of the hippocampus as a temporary memory store. *Phil Trans Roy Soc B* (in press)

Professor,
Division of Neural Systems, Memory and Aging,
University of Arizona, Tucson, Arizona 85724