united nations
educational, scientific
and cultural
organization

international atomic
energy agency

the
abdus salam
international centre for theoretical physics

SMR: 1343/14

EU ADVANCED COURSE IN COMPUTATIONAL NEUROSCIENCE

An IBRO Neuroscience School

(30 July - 24 August 2001)

"Evolution of Hippocampus and Neocortex"

presented by:

Alessandro TREVES

SISSA - International School for Advanced Studies
Settore di Neuroscienze Cognitive
Via Beirut 9
34014 Trieste
ITALY

HIPPOCAMPUS 6:666–674 (1996)

How Much of the Hippocampus Can Be Explained by Functional Constraints?

Alessandro Treves,[1] William E. Skaggs,[2] and Carol A. Barnes[2]

[1]SISSA, Cognitive Neuroscience, Trieste, Italy; [2]Neural Systems, Memory and Aging, Arizona Research Labs, Tucson, Arizona

ABSTRACT: In the spirit of Marr, we discuss an information-theoretic approach that derives, from the role of the hippocampus in memory, constraints on its anatomical and physiological structure. The observed structure is consistent with such constraints, and, further, we relate the quantitative arguments developed in earlier analytical studies to experimental measures extracted from neuronal recordings in the behaving rat.

KEY WORDS: associative memory, storage capacity, redundancy, forgetting, sparse coding

INTRODUCTION

This is not a computational model, really. Across scientific disciplines, *computational* usually qualifies approaches based on the use of the *computer*, as opposed to theoretical analysis or physical experiment. In the study of the brain, the term vaguely suggests, in addition, that an approach is aimed at the *computations* performed by a given ensemble of neurons or a given structure. What is reported in this article is related to the latter but not much to the former meaning of the word, since it is almost entirely based on formal analytical derivations, not on computer simulations. Moreover, the term *model* usually implies a definite system, as specified by a collection of formulas or by a set of computer instructions or even by an organic or living preparation, chosen to study in simplified form phenomena pertaining to the "original." Our approach is not based on the study of a definite model, but rather on the use of different formal models, each calibrated according to the specific questions asked, and on the model-independent analysis of neuronal activity.

The hippocampus is both a structure emerging from mammalian evolution and a system dedicated to its own particular operations on the information it processes. While some aspects of its organization may be the semi-accidental result of its evolutionary history, for others, in particular for the quantitative values of biologically tunable parameters, it is legitimate to argue that they must be geared to *optimize information processing.* This approach is aimed, then, at understanding which aspects of the organization of the hippocampus—anatomical or physiological—stem from this higher level requirement of optimizing the function it performs. At the most abstract level, this function is equivalent to manipulating information in certain ways; accordingly, information theory is the basis of our approach.

Knowledge about hippocampal anatomy and physiology are to be regarded, obviously, as an input to this approach. It is perhaps less obvious, but equally true, that ideas about the role which the hippocampus plays in managing information within the brain are also an input, and not an outcome, of this approach. In short, the goal of the approach is neither to discover structure nor to expound function, but solely to hypothesize or establish *explicit* relations, whenever possible, between structure and function. The success of the approach must be evaluated on the basis of the number of predictive (ideally quantitative) relationships it allows to be established, and the fraction of these that are validated by direct experiment. It should not be evaluated on its inability to explain those aspects of the structure which it does not link to function, nor on the inaccuracy, omissions, or fallacy of the structural and functional descriptions it builds upon.

THE HIPPOCAMPUS AS A MEMORY DEVICE

Marr's system level of the hippocampus (1971) was, in broad terms, the same description of its functional role in memory currently shared by several investigators and taken as an input for the present analysis. His perception of the role of formal models as providing explicit links between structure and function, leading to verifiable predictions (ranked with his curious star system), set the paradigm for others, including us, to follow. What was lacking in his time was 1) detailed knowl-

Accepted for publication July 15, 1996.
Address reprint requests to Alessandro Treves, SISSA, Cognitive Neuroscience, via Beirut 2-4, 34013 Trieste, Italy.

for two distinct input systems to the hippocampal CA3 network. Hippocampus 2:189–199.

Treves A, Rolls ET (1994) Computational analysis of the role of the hippocampus in memory. Hippocampus 4:374–391.

Willshaw D (1971) Models of distributed associative memory. Unpublished doctoral dissertation, University of Edinburgh.

Willshaw D (1981) Holography, associative memory, and inductive generalization. In: Parallel models of associative memory Hinton G, Anderson J, eds, pp 83–104. Hillsdale, NJ: Erlbaum.

Wilson MA, McNaughton BL (1994a) The preservation of temporal order in hippocampal memory reactivation during slow-wave sleep. Soc Neurosci Abstr 20:1206.

Wilson MA, McNaughton BL (1994b) Reactivation of hippocampal ensemble memories during sleep. Science 265:676–678.

Winocur G (1990) Anterograde and retrograde amnesia in rats with dorsal hippocampal or dorsomedial thalamic lesions. Behav Brain Res 38:145–154.

Zola-Morgan S, Squire LR (1990) The primate hippocampal formation: evidence for a time-limited role in memory storage. Science 250:288–290.

edge about both structure and function, but also 2) the mathematics adequate to analyze formal models refined enough for his purposes. Therefore 1) the discussion of how the theory would be implemented within the hippocampus remained rather vague, although it inspired subsequent work in which the correspondence was made more precise (McNaughton and Morris, 1987; Rolls, 1989), and 2) the quantitative results of his analysis were not applicable to the real system. Nevertheless, Marr's attempt to *explain* the hippocampus remains the most important reference point for later analyses.

Following Marr, the theory considered here for the function of the hippocampus is that it serves as an intermediate-term memory store, in which neural representations of certain events are stored on-line as the events are experienced, and from which they can be retrieved, off-line, by a so-called cue. This is a widespread conceptualization of the role played by the hippocampus, originally based on evidence from human patients (Scoville and Milner, 1957) and later discussed also in the context of experimental findings with other primates and rodents (see the debates following Rawlins, 1985; Eichenbaum et al., 1994). The findings largely agree that memory retention by the hippocampus is limited in time (Squire, 1992); what is more controversial is whether hippocampal forgetting follows the transfer—mediated by cued retrieval—of the same episodic information to neocortical permanent storage sites, possibly after reorganization into a semantic system (cf. Gaffan, 1993). In any case, typical forgetting or transfer times would for humans be of the order of years, whereas Marr (1971), who based a similar "transfer" notion on the idea that it would occur during sleep, when neocortex is shut off from sensory inputs, assumed these times to be of the order of days.

The main alternative theory, that the hippocampus operates as a spatial computer (e.g., O'Keefe, 1990), will not be considered here, not as a tacit denial that space has intimate connections with hippocampal function, but rather because we argue that the fact that the information being processed is wholly or partially spatial in nature does not necessarily constrain hippocampal structure (cf. Treves et al., 1992). Even within the "memory" camp, a substantial discussion has been devoted to the characterization of 1) the type of information which reaches the hippocampus and 2) the transformations it goes through within the hippocampus (e.g., Nadel, 1991, and the "forum" that follows). These are both important aspects, neither of which is attended to in this article. We focus only on simpler and more abstract quantitative aspects, such as the amount of information in a representation. For example, whether the spatial information in a representation is in terms of an egocentric or allocentric frame, or refers to the animal's location in the rat or to external space in the monkey, are possibilities which will not be discriminated here, as long as they correspond to *xx* bits of space information. If it were shown, to continue the example, that knowledge about where the rat is in the arena comes to the hippocampus in polar coordinates and is within the hippocampus transformed to Cartesian coordinates, this would again be irrelevant to the present discussion, except for the possible loss in resolution and information resulting from the transformation.

If we are not going to argue about what the hippocampus feeds itself with, nor about how it digests it, what is it that will matter to us? From our information-theoretical viewpoint, the hippocampus, in order to carry out the memory function it is specialized for, must be able to

1. Generate, on-line, appropriate neuronal representations of the events it has to store in memory;
2. Store these representations on-line, and thus in a single shot;
3. Hold multiple representations simultaneously in storage within the same system;
4. Retrieve each representation from partial cues;
5. Send back the retrieved information in a readable and robust format.

These simple requirements in fact significantly constrain the structure of the biological device that must fulfill them, especially if they are taken quantitatively, that is, if they have to be met with optimal or near-optimal solutions. This is what is discussed next, with a subsection devoted to each of the requirements, which, so as to follow the logic of the argument, are considered in the scrambled order 4, 2. 3, 5, 1: a content-addressable memory implemented as a cascade of Hebbian associative networks, with a free autoassociator at its core, a post-processor at the end, and a pre-processor at the front.

A Content-Addressable Memory . . .

Requirement 4, the ability to retrieve information from partial cues, that is from arbitrary subsets of the information to be retrieved, is equivalent to requiring that the device in question be a content-addressable memory. The qualitative, explicit nature of the cue could be further defined as being a sensory component of a multimodal episodic memory, or part of the context, or in many other ways. At a quantitative and abstract level, the utility of such a device arises from the difference between the amount of information that has to be supplied with the cue, and the amount of information that can be retrieved from the device. If this difference is zero or the former is more than the latter, the device does not operate as a memory but merely as a converter (although the activity of individual units within the device may still show memory effects, such as place cells maintaining their specificity in the dark). Therefore, quantifying the *information gain* provided by the system is crucial for establishing whether its role in memory is substantial or purely coincidental. The information content required for the cue to be effective in eliciting retrieval depends on the type of memory and on its load, in the same sense that an e-mail address has a different size, on average, from a regular mail address. If the memory is taken to hold p item (relevant ranges for p are discussed below), the minimum information necessary in the cue is

$$I_{\text{cue}} = \log_2 p. \qquad (1)$$

The information content of each memory item is also dependent on the type of memory, but in general for a system based on the parallel operation of N processors, if a memory item is represented by the values taken at a point in time by N variables associated

with each element, it may be expected to be of order

$$I_{item} \sim N\, i_{elem} \qquad (2)$$

where i_{elem} is the average information provided by individual elements taken separately.[1] Eq. 2, far from being a simple change of notation from I_{item} to i_{elem}, would express, if verified, a deep property of the type of representation used by hippocampal cells: its additivity in terms of information, or in other words that different cells convey different information. In contrast, if *redundant* representations were used, I_{item} would be much less than the sum of individual i_{elem}'s; whereas if the representations were *synergistic*, it would be more than the sum.

The most important condition, for our device to be effective as a memory, is then that I_{item} be much larger than I_{cue}, and this will be the case if Eq. 2 is approximately valid, that is if I_{item} grows linearly or almost linearly with N, and of course if N is large. Translated into plain hippocampal terms, the hippocampus can function as a memory if its many cells operate independently or near independently (a few global or quasi-global constraints are acceptable, whereas many detailed mutual constraints on their activity would be functionally destructive).[2]

A very important experimental result, then, is that the hippocampus indeed appears to display such functional diversity, in that the evidence available is consistent with Eq. 2. The collection of such evidence is not simple, because 1) it is only possible to extract the portion of the information contained in a pattern of cellular activations which is about a limited accessible set of correlates, such as position in space for hippocampal cells in the rat; 2) the bias in information estimates due to limited sampling (Treves and Panzeri, 1995) requires the collection of large amounts of recorded activity; and 3) recording from multiple cells simultaneously is a major task, which can now be handled but only up to about 100 cells (Wilson and McNaughton, 1993), still few compared with the hundreds of thousands present in the system. The first aspect, in particular, implies that regardless of whether the information provided by different cells is actually independent, the total measured information will never exceed a ceiling set by the log of the limited number of variables (here, spatial bins) considered.

Figure 1 shows that individual cells contribute independent information, at least until information values begin to saturate as they approach the ceiling of the maximum information possibly associated with spatial position, as defined in the experiment. In other words, different cells are *redundant* only inasmuch as they cannot answer more (about a definite question: here, spatial position) than a full answer (cf. Rolls et al., 1996). To a theoretical, infinitely complex question it is possible, and consistent with current evidence, that a population of N cells would provide an answer with an information content scaling with N, hence much larger than the minimum information required to be supplied with the cue (I_{cue}).

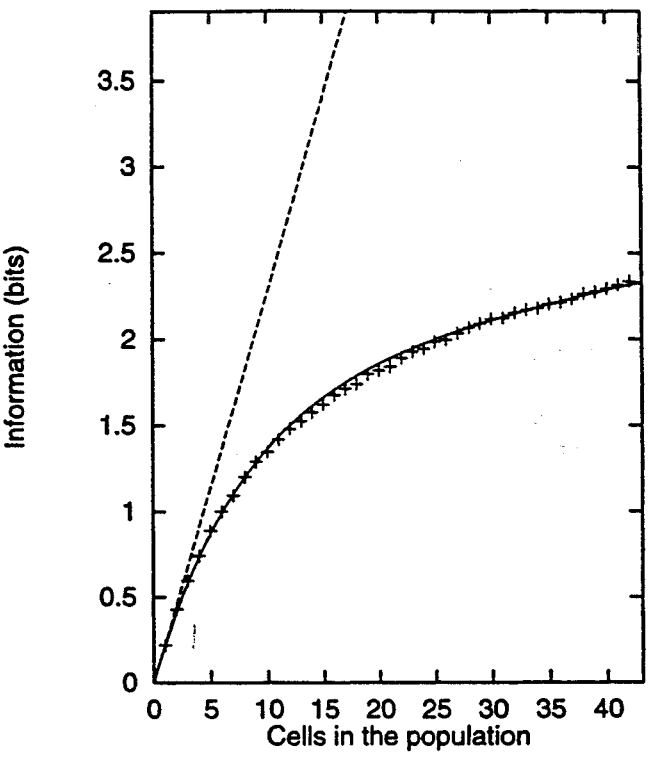

FIGURE 1. Average information extracted from subsets of hippocampal cells from a sample of 42, about the position of the rat in a three-arm maze, vs. the size of the subset. The data points (+) are fitted with a simple model (solid line) that explains the deviation from a linear increase solely in terms of ceiling effects: a separate ceiling $I^{arm}_{max} = \log_2 3$ for providing information about which arm the rat is occupying, and one $I^{segm}_{max} = \log_2 5$ for providing information about which of the five segments in which each arm has been discretized is currently occupied by the rat. The model is expressed by the equation

$$I(n) = I^{arm}_{max}[1 - \phi^n_{arm}] + I^{segm}_{max}[1 - \theta^n_{segm}], \qquad (3)$$

with the ϕ's fit parameters. The total ceiling $I_{max} = \log_2 15$ (15 being the number of spatial bins) is at the top of the ordinate axis. The linear increase extrapolated from the full information i_{elem} provided by the responses of single cells (not a fit parameter) is also indicated (dashed line). For details about this type of analysis see Rolls et al. (1996).

A separate issue to be examined is whether the theoretical minimum for I_{cue}, $\log_2 p$, is enough for a specific implementation of a content-addressable memory to effectively retrieve a memory item. This has been shown to be the case for the associative networks discussed below.

. . . Implemented as a Cascade of Hebbian Associative Networks . . .

Requirement 2, the ability to store neuronal representations in one shot, is satisfied by associative neuronal networks operating with *Hebbian* types of synaptic plasticity, as shown by a va-

[1] Essentially, the average difference between the total entropy of the variable associated with the element, and its entropy conditional to a given memory item.

[2] To focus on the common cohesive behavior of entire populations of cells, as often done in neurophysiology and in brain imaging (talking, e.g., about global neuromodulators, or epileptogenesis, or activated areas), is to negate the functionally useful aspects of parallel processing systems that arise from the diversity and incoherence of individual cell functions.

riety of formal models, including those considered by Willshaw et al. (1969), Kohonen (1977), and Hopfield (1982). Among all content-addressable memories, many others could be conceived, of course, that could store representations in one shot. We focus, however, on systems made up of neurons, characterized by a weighed summation of inputs from other units (as opposed to a completely arbitrary operation) and connected by synapses whose efficacy can be modulated by activity, but only locally in space and time. This restricts the class of systems with the required ability pretty much to variations of associative networks (Treves and Rolls, 1991). The essential reason for this is not the sophistication of associative networks, but precisely their simplicity. More complex networks (for example, backpropagation networks, in which, however, synaptic plasticity is not a local effect) tend to prefer iterative learning because their complexity yields to instabilities when faced with rapid, one-shot learning (McClelland et al., 1995). A basic associative memory function may nevertheless coexist, as a sort of minimum common denominator, with other functionalities within the same networks.

The crucial ingredient that endows networks of real neurons, operating with distributed representations, with an associative memory ability is a Hebbian type of synaptic plasticity, such as the type known as LTP (long-term potentiation), which is induced, e.g., through the action of *N*-methyl-D-aspartate (NMDA) receptors. Associative LTP is present at several synaptic systems in the hippocampus, including the most intensely studied perforant path to granule cells and Schaffer collateral to CA1 systems (but possibly not the mossy fiber to CA3 system). All such systems (not the mossy fibers) operating in cascade, along the direction of preferred activation flow, can contribute to an associative memory function.[3] The involvement of LTP in associative memory has never been conclusively demonstrated, nor it is likely that it will be in the near future (Barnes, 1995); nevertheless, its very existence provides what would otherwise be a missing link in our logic, and a serious one at that.

Turning to a quantitative analysis, the number p of representations that can be stored with any reasonably efficient type of Hebbian plasticity is broadly determined by the relation

$$p < p_c \sim 0.2 \frac{C}{a \log(1/a)} \qquad (4)$$

where C is the average number of associative inputs per cell, and a is the average sparseness (Treves, 1990) of the representations. This relation is valid for a large class of networks that store distributed representations (firing patterns) with Hebbian "learning rules" that model associative types of synaptic plasticity. It has been established originally for the sparsely coded version of the Hopfield autoassociator with binary units (Tsodyks and Feigel'man, 1988; Buhmann et al., 1989) and later found to hold also for a variety of specific models with graded response units that differ in the type of connectivity (the architecture), the statistics of the firing patterns, or the exact learning rule used (Treves and Rolls, 1991). It holds also for more realistic models that incorporate a description of the dynamics of real neurons (see below) and is thus expected to apply to real networks in the brain, even allowing for some semantic structure in the encoding of what the simplest models treat as independent episodic representations.[4]

The *sparseness* of the firing patterns, or intuitively speaking the proportion of units active in a representation, is the most important quantity that balances storage capacity with representational capacity. Sparse coding (a small) allows more memory items to be stored, but the information content of each item (hence the representational capacity of the network) decreases. Experimentally, relatively sparse coding is found to prevail in those areas such as the hippocampus (in rats, Barnes et al., 1990; and monkeys, Rolls et al., 1989) which are closely associated with a simple role in associative memory; whereas in areas whose role in memory is thought to be different, and in any case minor with respect to sensory encoding, such as the monkey temporal cortex, firing patterns are found not to be sparse at all (Rolls and Tovee, 1995).

If the quantities C and a vary across the cascade of networks, the relevant ones are of course those that produce the minimum p (that is, the memory bottleneck). Since C, whatever the type of connectivity, is of the order of N or less, p also cannot exceed N by much, and its logarithm, that is I_{cue}, is a small number with respect to I_{item} even if the mutual information per cell, i_{elem}, is a small fraction of a bit, as found to be the case with the sparse firing of hippocampal cells (see Fig. 1).

If p is limited and the memory device has to function over a lifetime, a need obviously arises for a mechanism that erases old memories and hence allows for the storage of new ones. Among the simplest of these *forgetting* mechanism (forgetting intended at the hippocampal level, not necessarily at the behavioral level) are constraints on the range of variability of synaptic efficacies—which are certainly present at least in the sense that synaptic conductances cannot become negative—and the gradual, passive decay of synaptic enhancement.[5] Denoting with τ_{item} the mean

[3]Plasticity in the mossy fiber system would be useless for associative memory purposes because of the small number of synapses per receiving cell (Treves and Rolls, 1992) but may be useful in satisfying requirement 1 (see below), although interestingly Kandel and coworkers (Huang et al., 1995) failed to find a learning deficit in rats following the selective blockade of such plasticity.

[4]Recent evidence (Skaggs and McNaughton, 1996) points at one basic "semantic" aspect that seems to be preserved in hippocampal memories: the temporal order in which different representations were activated. Besides, rat spatial maps can be regarded as a further semantic structure linking the memories of nearby positions in the environment.

[5]In the Marr (1971) model with binary synaptic elements, the limit p_c on p was set by the requirement that enough synapses remain unsaturated, that is at the lower of the two efficacy values. Thus, saturation determines storage capacity, and decay, which would be implemented as a "flip" back to the lower value, is the one mechanism for forgetting (McNaughton and Barnes, 1990). In the more efficient models that effectively use the formal equivalent of LTD (long-term depression) along that of LTP, storage capacity is determined by the balance between signal and noise, and saturation is an accessory ingredient which, if present, has an overwriting effect similar to (continuous) decay, thus constituting an independent potential mechanism for forgetting (Barnes et al., 1994).

permanence time of a representation in the hippocampal system, and with dp/dt the acquisition rate, we have $p \simeq \tau_{item} dp/dt$. If the hippocampus operates at close to its memory capacity, as an efficient system should, then

$$dp/dt \sim 0.2 \frac{C}{\tau_{item} a \log(1/a)} \quad (5)$$

which implies that if the acquisition rate varies strongly across time, some of the other three parameters should be tunable to maintain optimal performance. This could be verified experimentally by manipulating the acquisition rate and monitoring both sparseness and hippocampal forgetting through multiple single-unit recording and behavioral procedures, as suggested before (Treves and Rolls, 1994; Treves et al., 1996). Similarly, independent changes in C, for example a reduction with aging, might be partially compensated by tuning a or τ_{item} (Barnes et al., 1994). If, further, τ_{item} is related to the exponential time constant measurable in LTP decay, as their similar reduction with aging suggests (Barnes and McNaughton, 1985), this might allow us to follow changes in the time parameter with purely neurophysiological means. These aspects might also be probed, in the future, by direct measures of the average plasticity in vivo, as explained in detail elsewhere (Treves et al., 1996).

. . . with a Free Autoassociator at its Core . . .

We now examine requirement 3, considering how to optimize storage capacity by choosing the most suitable network *architecture*, once the elements available are given (pyramidal cells, NMDA receptors, etc.), and the sparseness is set as yielding the best balance with the information content of each memory. A free autoassociator, that is an autoassociative memory concentrated in a compact network heavily interconnected with recurrent collaterals (Treves and Rolls, 1991), would be the most efficient link in the posited associative chain, on at least two accounts: the ease with which it can store new representations and the full use it would make of memory space. The first factor arises from each output cell having access, in such an architecture, to both afferent inputs and, through at most a few synaptic steps, collateral inputs from all the other cells in the net. This will be discussed further below. The second factor arises from I_{item} being proportional to N, as discussed above, with this N being in a single-layer network both the total number of cells *and* the number of output cells. Since p is proportional to C, and the *total information* the net stores at a moment in time is just pI_{item}, this information is proportional to CN, which for a free autoassociator is also the number of synapses in the network. In fact, a reasonable estimate, established analyzing a wide class of formal models, is that 0.1–0.2 bits could be stored per synapse. Directed associative memory networks (Marr, 1971), instead, are either monolayer, and then unable to produce complete retrieval (Gardner-Medwin, 1976), or multilayer, and then the output cells are just a subset of the cells in the network, as more cells are present that contribute to memory retrieval but do not access stations downstream. Although these nets can still function as autoassociators (Treves and Rolls, 1991), they use the available synaptic space less efficiently. We believe that this simple information-theoretic advantage has provided most of the evolutionary pressure for the emergence of the extensive CA3 recurrent collateral system. The value of C, about 12,000 in the rat (Amaral et al., 1990), could be interpreted as having been maximized, in order to increase storage capacity, under the constraint of maintaining cells electrically compact, to ensure that Hebbian plasticity at the dendrites reflects events occurring at the soma. Dendrites operating as effectively independent units (Softky, 1994) would not implement an associative memory.

A recurrent autoassociator can be seen as iterating in time the same operation performed just once in an equivalent but purely feedforward system. This implies *feedback*, which has several effects, and could also be taken to imply a very long time to operate, involving, as it were, repetitive cycles through the recurrent circuit. The second expectation is borne out of considering overly simplified nondynamical models, but it is contradicted by the analysis of formal models that include the relevant dynamical biophysics of pyramidal neurons (Treves, 1993). This analysis, corroborated by computer simulations, shows that recurrent collaterals can contribute their effect over short times, of the order of the time constant for conductance inactivation at their synapses. Feedback, on the other hand, results in the following: 1) it complicates considerably the analytical methods required to understand the properties of formal models (Amit, 1989); 2) if strong, it allows for self-sustained activation, which can subserve short-term memory; 3) it amplifies interference among different memory representations, but very little if the coding is sparse (Treves and Rolls, 1991), as it appears to be in the hippocampus; 4) it can make it difficult for subtractive inhibition to control the activity of the pyramidal cells, while at the same time allowing them to operate efficiently as a memory—solving this conflict requires shunting inhibition (Battaglia and Treves, 1996). Effect 1 is worrisome for the modelers but not for the hippocampus; effect 2 may or may not be used by the hippocampus; effect 3 is effectively avoided by sparse coding, which is already there for other reasons (see above); effect 4 merely puts additional constraints on the organization of the inhibitory component of the circuitry, and disruption of these constraints accounts for the ease with which hippocampal activity can get out of hand, e.g., in epilepsy (Traub and Miles, 1991). The existence of CA3 with its prominent recurrent collateral network suggests that these side effects are overridden by the two advantages that a compact associative net provides, in terms of learning and in terms of effective usage of synapses.

. . . a Post-Processor at the End . . .

The information retrieved within the device, in particular by the autoassociator, has to be sent back to "external users," i.e., neocortical areas, with minimal waste—requirement 5. This can be accomplished by a multilayered system of Hebbian-modifiable backprojecting synapses (see Treves and Rolls, 1994). In addition, the CA1 network can be considered to be both the first step in the relay from CA3 back to neocortex and the last stage of the hippocampal associative memory system, in other words a dedicated memory post-processor. One crucial advantage of having

CA1 after the CA3 stage, is that the very compressed representation provided by CA3 pyramidal cells (in numbers, the bottleneck of the hippocampal system) can be reexpanded onto the larger number of CA1 pyramidal cells. The reexpansion appears to occur across species (Seress, 1988) and results in the same information being coded in a much more robust manner. As an artificial but intuitive example, if N CA3 cells code for $2N$ bits of information by firing at 1 of 4 equiprobable rates, $2N$ CA1 cells can code the same information one bit each, by firing at 1 of just 2 equiprobable levels, with a corresponding increase in the permissible noise level. Obviously the recoding is effective only if at least it *preserves* the overall information content of the representation. Further, CA1 can also contribute to associative retrieval itself by *increasing* this information content over and beyond that of the representation retrieved from CA3. A quantitative assessment of such information content using an analytical model (Treves, 1995) shows that both preservation and increase can occur if the CA3-to-CA1 connections, the Schaffer collaterals, are endowed with associative Hebbian modifiability. In particular, there is an optimal range of the plasticity parameter (measuring essentially the average modification per item as a fraction of the total variance) which is the one that matches the plasticity of CA3 recurrent connections.

The analysis then suggest that the two synaptic systems (within CA3 and from CA3 to CA1) may be optimally organized if they share the same type of synaptic plasticity, in particular the same molecular and biophysical mechanism based on NMDA receptors. This analysis is now being extended, through an appropriately adapted formalism, to include the direct perforant path projections to CA1 (Panzeri, Fulvi Mari, and Treves, in preparation). It is hoped that this will clarify the contribution of direct entorhinal inputs to the information content in the hippocampal output and provide indications as to the reason for the relative abundance of perforant path and Schaffer collateral synapses onto CA1 cells.

If no substantial increase is contributed by direct perforant inputs, the information *per cell*, I_{item}/N_{CA1}, provided by a population of CA1 cells should then be lower than in CA3 in the inverse ratio of the number of cells, N_{CA3}/N_{CA1}. This is what is found in preliminary analyses of data recorded simultaneously in CA3 and CA1, and kindly provided by Matt Wilson. The ratio is about 1.4 in the rat, and the analysis must select homogeneous samples, with similar single-cell firing statistics in the two areas (implying similar values for i_{elem}). In other words, the representation provided by CA1 cells appears, from preliminary evidence, to be slightly more redundant, and hence more robustly coded, than in CA3.

Note that, in addition, firing patterns appear to be sparser in CA3 than in CA1, and mean rates lower (Barnes et al., 1990). The more pronounced sparseness may be related to CA3 being the recurrent autoassociator, hence more subject to interference. The lower mean rates are in principle a separate issue (exactly the same sparseness would be measured if all firing rates were scaled down by a common factor), but slow firing may be related to sparse firing, at least at a general hippocampal level, in the following indirect sense. The distribution, across, e.g., spatial positions, of quasi-stationary currents entering the soma has a more or less fixed shape for any cell receiving many small distributed inputs (quasi-Gaussian, in fact); these currents result in a distribution of firing rates through basically a threshold process, and one of the simplest ways to modulate the sparseness of the distribution of rates is to alter thresholds: High thresholds yield sparser distributions than low ones (see, e.g., Treves and Rolls, 1992). Therefore the setting of relatively high thresholds, with the side effect of low mean firing rates, may be a mechanism aimed, in the hippocampus, toward sparser codings.

. . . and a Pre-Processor at the Front.

Finally, the first requirement in our list is that adequate mechanisms exist to generate appropriate representations for memory storage. Appropriate, in the abstract sense considered here, means rich in information about the event that is being represented—as rich as is compatible with the compression inherent in using a compact associative network, and with the sparseness required to store a large number of representations. Treves and Rolls (1992) have produced a quantitative argument that indicates the advantage of delegating this task to a specialized system, the dentate gyrus, with its sparse mossy fiber projections onto CA3 cells. A different argument suggests instead that the direct perforant path to CA3 is the system apt to relay to CA3 cells the cue that initiates the retrieval of a representation. For the mossy fiber inputs to provide the required driving effect on the firing of CA3 cells, overcoming the interference resulting from the activation of recurrent collaterals, the mossy fibers need not be quite as strong and specific as *dentonators* (cf. McNaughton and Morris, 1987): The observed physiological and anatomical properties would be adequate.

A direct experimental check of the theoretical argument entails the quantification of the information contents of CA3 representations learned before and after the inactivation or ablation of the granule cells. In intact rats, the observed values of information per cell as a function of sparseness are consistent with expectations based on formal models, as shown in Figure 2.

Following the near-complete destruction of granule cells, the corresponding values should go below the lower curve of Figure 2. This would be a quantitatively sensitive way to probe an effect that may be unclear at the behavioral level. Analysis of data recorded from lesioned rats is in progress and will be reported elsewhere.

The dentate gyrus, if this hypothesis is found to be correct, might thus be regarded as a late addition to the hippocampal system (granule cells have a late ontogenetic development) that serves to greatly increase the information-theoretic efficiency of its associative networks.

The curves in Figure 2 are derived from simple formal models, in which the relevant distribution of firing rates is taken to be the asymptotic limit for long times, whereas the data points, being derived from actual experiments, must correspond to the measurement of firing rates as spike counts over a finite time interval. Experimental evidence, however, indicates that for a freely running rat the distribution of mean rates at each spatial position

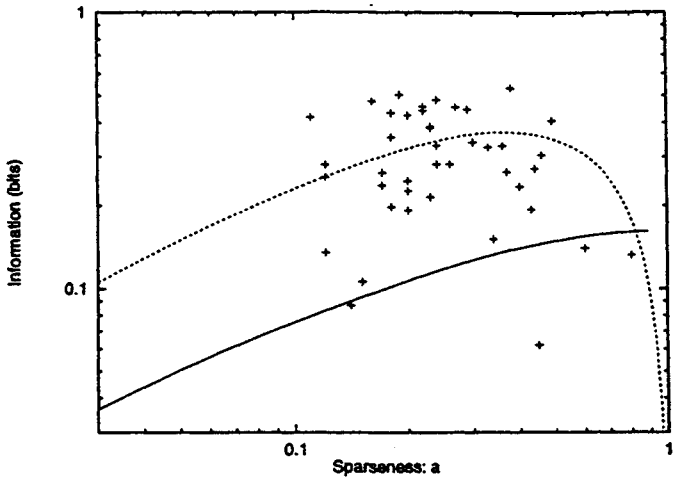

FIGURE 2. **The information extracted from single hippocampal cells from a sample of 42, about the position of the rat in a three-arm maze, vs. the sparseness of their firing. Given the number of spatial bins, the sparseness can only range from 1/15 = 0.067 to 1. The upper curve is the expected average trend of $i_{elem}(a) = a\ln(1/a)$ (Treves and Rolls, 1992), while the lower curve is the expected *upper* limit of the data points if mossy fiber inputs were absent in the storage of new representations (see Treves and Rolls, 1992).**

(and with it the sparseness of the distribution) does not vary much with the size of the time bin over which they accumulate; whereas the variance of such rates around their means shrinks, as of course it should, with longer bins, tending to a finite non-zero value which represents the intrinsic variability in the rates. Correspondingly, the information extracted from spike counts of increasing bin length rises until it saturates at bin sizes of order of 1 s or less, and this can be taken to be the value that should match the asymptotic values of formal models. The instantaneous information rates (Skaggs et al., 1993), instead, depend only on the distribution of mean rates and are relatively constant with the size of the time bin. These facts are illustrated in Figure 3, which exemplifies single-cell information kinetics in the hippocampus.

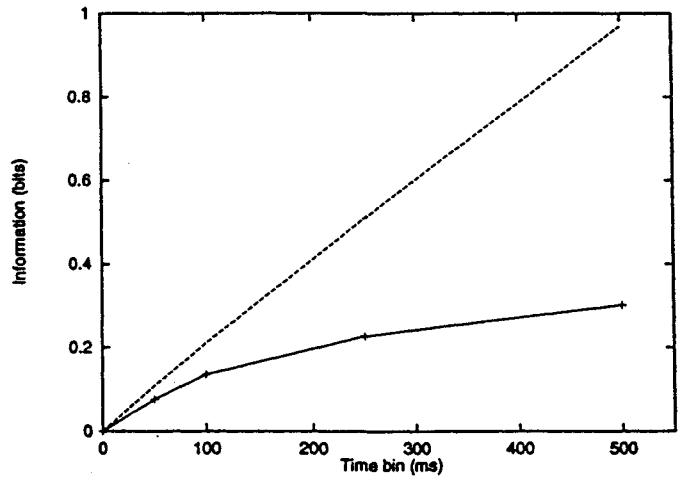

FIGURE 3. **Average information extracted from single hippocampal cells from a sample of 42, about the position of the rat in a three-arm maze, vs. the size of the time bins used for counting spikes: 50, 100, 250, or 500 ms. For comparison, the information values, obtained by multiplying the instantaneous rates by the corresponding time bin, are also displayed by the dashed curve, which being nearly straight implies almost constant instantaneous information rates.**

DISCUSSION

What Is New in This Approach?

The approach itself is not new, because it has been evolving for 25 years, but new powerful analytical methods have emerged for understanding in more detail both formal models of associative memories and data recorded from hippocampal cells. The latter are now obtained with massively parallel simultaneous recordings, which opens up the possibility of analyzing the information conveyed by populations instead of just single cells. As a result of these developments, the correspondence between the requirements of the theory and the actual structure of the hippocampus is improving, as documented in this article.

What Good Is the Math?

The arguments recapitulated above are quantitative, and as such they have to rely on the conjunct mathematical analysis of both adequate formal models and recordings of neuronal activity in vivo. A quantitative analysis is crucial to understanding the relationship between structure and function in the hippocampus, because an entirely different structure may subserve the same qualitative function, at the expense of information-theoretic efficiency. For example, a structure without the dentate or even without CA3 can still function as an intermediate-term memory that stores and retrieves representations from partial cues. Thus, it would not be surprising if behavioral studies were to find limited impairments following complete CA3 ablation. The avian structure that is supposed to be an analogue of the hippocampal formation and is implicated in spatial memory (Krebs et al., 1989) does not show, as far as the anatomy is known, anything like the same internal structure, yet it may well take a similar functional role, with different efficiency.

A point of central importance, reflected in the suggestions for experiments of the last section, is that behaviour alone, as observable from outside the "black box," is certainly what ultimately matters, but is not a probe sensitive enough to really bear on the relations between structure and function within the hippocampus. Only the behavior *of the hippocampus*, as observed by recording the activity of its units, can inform a detailed understanding of the hippocampal formation. This cellular behavior has, however, to be observed in vivo, and of course the concurrent whole animal behavior is a most useful auxiliary variable to keep track of. On a similar note, stressing the crucial role played by the mathematical analysis of formal models does not mean denying the utility of computer simulations, often important especially in the preliminary investigation of questions not yet reduced to a formulation accessible to analytical methods.

What Is Different From Other Approaches?

Several other researchers share the same or a similar system-level view of the role of the hippocampus in memory but differ in the particular aspects they address or emphasize in their analysis. This should be evident also from reading other contributions to this issue. Rolls (this issue), for example, has instantiated this view in a real *computational* approach, by implementing a computer simulation of the operation of the hippocampus and entorhinal cortex, along essentially the same lines followed in our arguments. Murre (this issue), with his TraceLink computer model, does not attempt to account for the details of hippocampal circuitry, but rather for the phenomenology of amnesia in humans; but the basic ideas are, again, consistent with ours. In particular, Murre assumes that the fundamental constraint preventing on-line learning of episodic information directly in neocortex, thus generating the need for a dedicated hippocampal system, is the limited long-range cortico-cortical connectivity. In contrast, McClelland (this issue) has suggested that the fundamental constraint is that neocortical memory systems could only accommodate slow learning—and hence would require an auxiliary hippocampal fast system—if they operated like certain connectionist models.

The work of Hasselmo and colleagues (1995; and also reported in this issue), on the other hand, although consonant in perspective, focuses specifically on cholinergic modulation of the hippocampus, and proposes a mechanism which may be alternative to that proposed by Treves and Rolls (1992). Based on the observation that acetylcholine suppresses transmission by intrinsic connections and increases cellular excitability and synaptic modifiability, it is proposed that the switching "on" and "off" of the cholinergic input may be sufficient to differentiate between a storage and a retrieval phase, in particular by suppressing, when "on," transmission by CA3 recurrent collaterals, with its interference on the storage of new memories. One should note that the Treves and Rolls proposal does require a mechanism that switches off or attenuates *mossy fiber* inputs to CA3 during the *retrieval* phase; this may result from a decrease in the firing of dentate granule cells, but may also be helped by cholinergic action. However the Hasselmo proposal may be alternative in that, if cholinergic effects suffice in strongly suppressing *recurrent collateral* transmission during *storage*, no need would arise for separate input systems to CA3.

What Is There To Do?

There are many experiments that would greatly improve our understanding of the hippocampus from the viewpoint discussed here. The most important directions to take are as follows:

- **Parallel recording from monkeys.** While the considerations above have been mainly sculpted by data recorded in the rat, the rat or even rodent hippocampus is clearly not necessarily representative of the mammalian—and even less of the primate—one. Important new aspects have emerged from single-unit recordings in monkeys (O'Mara et al., 1994; Ono et al., 1993) and now actively walking monkeys (Rolls et al., 1995), whose implications will be even more crucial once the activity of populations will be analyzed.
- **Gradient of retrograde amnesia.** This has largely been a moot issue (Gaffan, 1993) because its exploration at the behavioral level confounds hippocampal forgetting with that resulting from other factors. An analysis of forgetting at the cellular level, with chronic implants, but also at the population level, without, will clarify 1) its existence and 2) its potential modulation by acquisition rate, aging, etc. (Barnes et al., 1994; Treves et al., 1996).
- **Differences between processing stages.** The comparison between data recorded from different populations (Barnes et al., 1990) has already been very insightful, but more detailed analysis of parallel recordings is needed to quantify the informational properties of each component of the whole system.
- **Selective blockades of plasticity.** Along with selective lesions, the disabling of plasticity at individual synaptic systems, with the fast developing genetic means (e.g., Huang et al., 1995), but even better with nonpermanent pharmacological means (Barnes, 1995), may allow us to verify or disprove the detailed predictions arising from this approach (see Treves and Rolls, 1992, 1994).

Acknowledgments

Different parts of the work described here were in collaborations with Edmund Rolls, Bruce McNaughton, Stefano Panzeri, and Francesco Battaglia, all of whom are to be thanked. Partial support came from grants AG12609 and MH01227 (USA), CNR94.02931.CT04 (Italy) and ERB-CHRX-CT93-0245 (EC).

REFERENCES

Amaral DG, Ishizuka N, Claiborne B (1990) Neurons, numbers and the hippocampal network. Prog Brain Res 83:1–11.

Amit DJ (1989) Modelling Brain Function. Cambridge Univ Press, New York.

Barnes CA (1995) Involvement of LTP in memory: Are we "searching under the street light?" Neuron 15:751–754.

Barnes CA, McNaughton BL (1985) An age comparison of the rates of acquisition and forgetting of spatial information in relation to long-term enhancement of hippocampal synapses. Behav Neurosci 99: 1040–1048.

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.

Barnes CA, Treves A, Rao G, Shen J (1994) Electrophysiological markers of cognitive aging: region specificity and computational consequences. Semin Neurosci 6:359–367.

Battaglia FP, Treves A (1996) Information dynamics in associative memories with spiking neurons. Soc Neurosci Abstr 22:445.4.

Buhmann J, Divko R, Schulten K (1989) Associative memory with high information content. Phys Rev A 39:2689–2692.

Eichenbaum H, Otto T, Cohen NJ (1994) Two functional components of the hippocampal memory system. Behav Brain Sci 17:449–472.

Gaffan D (1993) Additive effects of forgetting and fornix transection in the temporal gradient of retrograde amnesia. Neuropsychologia 31: 1055–1066.

Gardner-Medwin AR (1976) The recall of events through the learning of associations between their parts. Proc R Soc Lond B 194:375–402.

Hasselmo ME, Schnell E, Barkai E (1995) Learning and recall at excitatory recurrent synapses and cholinergic modulation in hippocampal region CA3. J Neurosci 15:5249–5262.

Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA 79:2554–2558.

Huang Y-T, Kandel ER, Varshavsky L, Brandon EP, Qi M, Idzerda RL, McNight GS, Bourtchouladze R (1995) A genetic test of the effects of mutations in PKA on mossy fiber LTP and its relation to spatial and contextual learning. Cell 83:1211–1222.

Krebs JR, Sherry DF, Healy SD, Perry VH, Vaccarino AL (1989) Hippocampal specialization of food storing birds. Proc Natl Acad Sci USA 86:1388–1392.

Kohonen T (1977) Associative memory. Berlin: Springer.

Marr D (1971) Simple memory: a theory for archicortex. Philos Trans R Soc Lond B 262:24–81.

McClelland JL, McNaughton BL, O'Reilly RC (1995) Why there are complementary learning systems in hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol Rev 102:419–457.

McNaughton BL, Barnes CA (1990) From cooperative synaptic enhancement to associative memory: Bridging the abyss. Semin Neurosci 2:403–416.

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

Nadel L (1991) The hippocampus and space revisited. Hippocampus 1:221–229.

O'Keefe J (1990) A computational theory of the hippocampal cognitive map. Prog Brain Res 83:301–312.

O'Mara SM, Rolls ET, Berthoz A, Kesner RP (1994) Neurons responding to whole-body motion in the primate hippocampus. J Neurosci 14:6511–6523.

Ono T, Nakamura K, Nishijo H, Eifuku S (1993) Monkey hippocampal neurons related to spatial and nonspatial functions. J Neurophysiol 70:1516–1529.

Rawlins JNP (1985) Associations across time: the hippocampus as a temporary memory store. Behav Brain Sci 8:479–496.

Rolls ET (1989) Functions of neuronal networks in the hippocampus and neocortex in memory. In: Neural models of plasticity (Byrne JH, Berry WO, eds), pp 240–265. San Diego: Academic Press.

Rolls ET, Tovée MJ (1995) Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. J Neurophysiol 73:713–726.

Rolls ET, Miyashita Y, Cahusac PMB, Kesner RP, Niki H, Feigenbaum J, Bach L (1989) Hippocampal neurons in the monkey with activity related to the place in which a stimulus is shown. J Neurosci 9:1835–1845.

Rolls ET, Robertson RG, Georges-Francois P (1995) The representation of space in the primate hippocampus. Soc Neurosci Abstr 21:586.10, 1494.

Rolls ET, Treves A, Tovée MJ (1996) The representational capacity of the distributed encoding of information provided by populations of neurons in the primate temporal visual cortex. Exp Brain Res, in press.

Scoville WB, Milner B (1957) Loss of recent memory after bilateral hippocampal lesions. J Neurol Neurosurg Psychiatry 20:11–21.

Seress L (1988) Interspecies comparison of the hippocampal formation shows increased emphasis on the regiosuperior in the Ammon's horn of the human brain. J Hirnforsch 29:335–340.

Skaggs WE, McNaughton BL (1996) Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. Science 271:1870–1873.

Skaggs WE, McNaughton BL, Gothard KM, Markus EJ (1993) An information-theoretic approach to deciphering the hippocampal code. In: Advances in neural information processing systems 5 (Hanson SJ, Cowan JD, Giles CL, eds). pp 1030–1037. San Mateo: Morgan Kaufmann.

Softky W (1994) Sub-millisecond coincidence detection in active dendritic trees. Neuroscience 58:13–41.

Squire LR (1992) Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. Psychol Rev 99:195–231.

Traub RD, Miles R (1991) Neuronal networks of the hippocampus. New York: Cambridge Univ Press.

Treves A (1990) Graded-response neurons and information encodings in autoassociative memories. Phys Rev A 42:2418–2430.

Treves A (1993) Mean-field analysis of neuronal spike dynamics. Network 4:259–284.

Treves A (1995) Quantitative estimate of the information relayed by the Schaffer collaterals. J Comput Neurosci 2:259–272.

Treves A, Panzeri S (1995) The upward bias in measures of information derived from limited data samples. Neural Comp 7:399–407.

Treves A, Rolls ET (1991) What determines the capacity of autoassociative memories in the brain? Network 2:371–397.

Treves A, Rolls ET (1992) Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. Hippocampus 2:189–199.

Treves A, Rolls ET (1994) Computational analysis of the role of the hippocampus in memory. Hippocampus 4:374–391.

Treves A, Miglino O, Parisi D (1992) Rats, nets, maps and the emergence of place cells. Psychobiology 20:1–8.

Treves A, Barnes CA. Rolls ET (1996) Quantitative analysis of network models and of hippocampal data. In: Perception, memory and emotion: Frontier in neuroscience (Ono T, McNaughton BL, Molotchnikoff S, Rolls ET, Nishijo H, eds). in press. Oxford: Elsevier.

Tsodyks MV, Feigel'man MV (1988) The enhanced storage capacity in neural networks with low activity level. Europhys Lett 6:101–105.

Willshaw DJ, Buneman OP, Longuet-Higgins HC (1969) Non-holographic associative memory. Nature 222:960–962.

Wilson M, McNaughton BL (1993) Dynamics of the hippocampal ensemble code for space. Science 261:1055–1058.

Neurobiology of Learning and Memory 73, 243–257 (2000)
doi:10.1006/nlme.1999.3931, available online at http://www.idealibrary.com on IDEAL®

Reversible Inactivation of the Hippocampal Mossy Fiber Synapses in Mice Impairs Spatial Learning, but neither Consolidation nor Memory Retrieval, in the Morris Navigation Task

Jean-Michel Lassalle, Thierry Bataille, and Hélène Halley

Laboratoire d'Ethologie et Psychologie Animale, UMR 5550 CNRS, Université Paul Sabatier, Toulouse, France

The role played by hippocampal mossy fibers in the learning and memory processes implemented in the Morris swimming navigation task has been studied in C57BL/6 mice by selective and reversible inactivation of mossy fiber synaptic fields by diethyldithiocarbamate. The functional integrity of the mossy fibers proved essential for the storage of the spatial representation on the modifiable synapses of the recurrent collaterals of the CA3 pyramidal cells, whereas it is not necessary for the consolidation and recall of spatial memories. The results suggest that mossy fibers are preferentially involved in new learning. They are consistent with the hypothesis that the hippocampal CA3 region might act as an autoassociation memory.

INTRODUCTION

One of the most crucial problems that animals have to solve when they live in the wild is that of spatial orientation. They have to acquire and memorize information from spatial cues and beacons to be able to orient themselves toward an invisible goal or to make a shortcut retreat toward a shelter.

Lesion studies have emphasized the prominent role played by the hippocampus in the processing of spatial information (Sutherland & Rudy, 1988; O'Keefe, 1991; Morris, Garrud, Rawlins, & O'Keefe, 1992). Unfortunately, permanent lesions do not allow conclusions to be drawn as to the specificity of the role played by the hippocampus because all the various stages of the learning and memory processes are affected. Reversible lesions, on the other hand, allow a refined interpretation of the brain mechanisms involved in

This research was supported by a grant from the Fondation pour la Recherche Médicale to J.-M. Lassalle. The authors gratefully acknowledge Paul E. Gold, Pascal Roullet, and two anonymous reviewers for helpful comments and suggestions.

Correspondence and reprint requests concerning this article should be addressed to Jean-Michel Lassalle, Laboratoire d'Ethologie et Psychologie Animale, UMR 5550, Université Paul Sabatier, Bat IVR3, 118 route de Narbonne, 31062 Toulouse cedex 04, France. Fax: 33 5 61 55 61 54. E-mail: lassalle@cict.fr.

1074-7427/00 $35.00

behavior, as underlined by Bures and Buresova (1990). For instance, they make possible the dissociation of the effects of the structural lesion on the performance vs. the learning process. With reversible lesions, the subject in its normal state can be tested again after the effect of the temporary deafferentation has disappeared. It can thus be used as its own control so that acquisition, consolidation, or memory recall can be evaluated independently in the absence of the target structure. Thus, Gallo and Candida (1995) showed that the reversible inactivation of the dorsal hippocampus by tetrodotoxin selectively impairs acquisition but not retrieval of the conditional blocking of taste aversion in rats. Electrophysiology studies realized in the early seventies by O'Keefe and Dostrovsky (1971), then more recently by O'Keefe and Burgess (1996), Burgess and O'Keefe (1996), and Cressant, Muller, and Poucet (1997), have shown that, in the CA1 and CA3 regions of the rodent hippocampus, there are place cells which respond to the spatial location of the subject. Cells which respond to the orientation of the head have also been discovered in the postsubiculum (Taube, Muller, & Rank, 1990) and in various other brain structures (Blair & Sharp, 1996). They probably make up the basic elements of a neural net that provides the animal with a spatial representation of its vital domain (Dudchenko & Taube, 1997).

Behavioral neurogenetics has shown that the variation of structures in the brain is controlled by genetic factors (see Lassalle, 1996, for a review) and led us to question their internal functioning, namely, to try to understand how the genetic variation of the hippocampal circuitry can control cognitive abilities. Different results have shown, sometimes with conflicting evidence, that the intraspecific variations in the size of the different hippocampal mossy fiber synaptic fields present genetic correlations with the variation of novelty responses (Crusio, Schwegler, & Van Abeelen, 1989; Roullet & Lassalle, 1990), open-field activity (Hausheer-Zarmakupi et al., 1996), intermale aggression (Guillot et al., 1994), and with various forms of associative (Lipp, Schwegler, Crusio, Wolfer, Leisinger-Trigona, Heimrich, & Driscoll, 1989) and spatial learning (Schwegler, Crusio, Lipp, Brust, & Mueller, 1991; Schwegler & Crusio, 1995). Our aim was to analyze the role played by mossy fibers in hippocampal functioning and, if possible, to find clues that would allow to understand through what kind of mechanism the variation in size of the mossy fiber synaptic fields could influence behavioral differences between strains of mice.

Treves and Rolls (1992, 1994) and Rolls (1994) proposed a functional hypothesis of the role played by mossy fibers. Their model assigns the hippocampal circuitry of the CA3 region (see Fig. 1) the role of an autoassociation network that would allow the storage of neural representations of episodic memories or spatial representations, the recall of which can be triggered by a fragmental input (see also McNaughton & Smolensky, 1991). This model predicts that mossy fiber synapses are essential to drive information storage, which corresponds to the process of learning. On the other hand, they are not necessary for memory retrieval, which is initiated by the synapses of the alvear pathway. The aim of the present work is to test this model. In order to dissociate the role played by the mossy fibers in the learning and memory processes of a spatial location in the Morris navigation task, we used selective and reversible inactivation of hippocampal mossy fiber synapses. This was obtained by diethyldithiocarbamate (DDC) infusions, a powerful technique the interest in which in behavioral studies remains nevertheless unexplored. DDC chelates the zinc contained in the giant mossy fiber synapses (Haug, 1967;

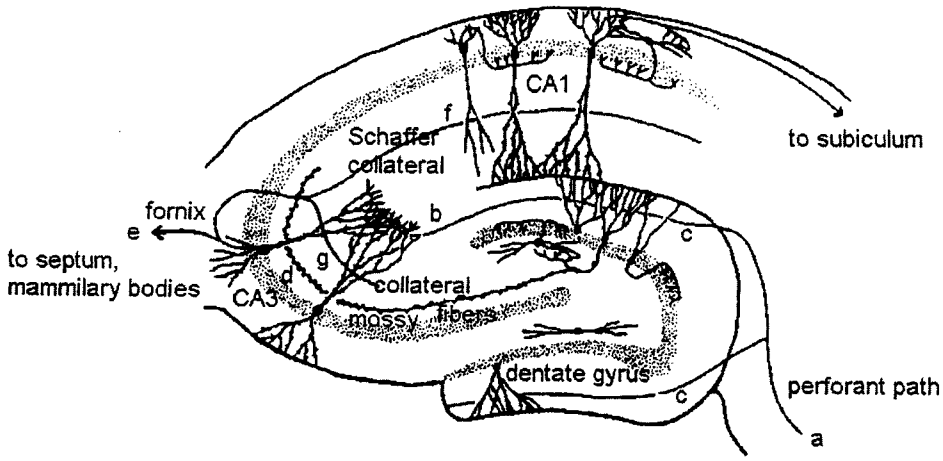

FIG. 1. Schema of the intrahippocampal connections (adapted from Rolls, 1994). The inputs from the entorhinal cortex reach the hippocampus through the perforant path (a). Some of these afferences synapse directly on the distal part of the apical dendrites of the CA3 cells (b) and constitute the alvear path, whereas the other perforant axons relay on the granular cells of the dentate gyrus (c). The axons of the granular cells contact "en passant" the proximal part of the apical dendrites and the basal dendrites of the CA3 cells, close to the cellular body, through the giant mossy fiber synapses (d). Three collaterals issue from the CA3 axon. One projects to the lateral septum and mammilary bodies through the fornix (e). The Schaffer collateral innervates the apical dendrites of the CA1 pyramidal neurons (f). The associative recurrent collateral contacts the median part of the apical dendrites of neighboring CA3 neurons by modifiable synapses (g).

Perez-Clausell & Danscher, 1985) which are inactivated for a 30- to 45-min duration, resulting in reversible working memory disruption (Frederickson, Frederickson, & Danscher, 1990). This time interval is long enough to allow a three-trial learning session in the Morris navigation task to be performed. Under the same conditions, control mice received an infusion of Ca-ethylenediamine tetraacetic acid (Ca-EDTA), which is also a zinc chelator that cannot penetrate the synaptic membrane and does not chelate the zinc within axonal boutons (Fredens & Danscher, 1973). This constitutes the right control for such an experiment. Infusions were made in the dorsal hippocampus which appeared the most appropriate location for that. Indeed, an accumulating body of evidence suggests that the hippocampus is a functionally and genetically heterogeneous structure along its rostrocaudal axis. For instance, Wimer and Wimer (1985) claimed that the hippocampus is a highly differentiated structure dependent upon four different genetic systems. More recently, Moser et al. (1993) have shown, in female rats, that the ventral and dorsal parts of the hippocampus may process qualitatively different kinds of information, the dorsal part being more important for spatial learning than its ventral counterpart. They showed that a 20% lesion volume of the dorsal hippocampus was sufficient to produce a long-lasting deficit in spatial learning in the Morris navigation task, whereas to be effective, a ventral lesion had to be large enough to have lesioned some cells of the dorsal hippocampus. Consequently, it appeared more relevant to study the effects of focal lesions of the mossy fiber pathway in the dorsal hippocampus.

Two experiments were carried out. The first was designed to dissociate the effects of the inactivation of the mossy fibers by DDC in the acquisition and memory processes of the spatial task and to assert the reversibility of the effects of DDC. The second was

planned to analyze the effects of DDC on the processes of memory consolidation and recall and to replicate the results of the first experiment.

GENERAL METHODS

Subjects

Ninety- to 120-day-old male mice from the C57BL/6 inbred strain were used. All mice came from the IFFA-CREDO breeding center and were 6–7 weeks old at their arrival at the laboratory. They were housed by groups of five to six in 30 × 20 × 14 polycarbonate cages placed in a rearing room at constant temperature (23 ± 1 °C) with a reversed 12–12 LD schedule, the onset of the dark phase being at 8:30 AM. Food and water were given ad libitum. Sawdust bedding was changed only once a week, at the end of each series of experiments. Experiments were always run in the afternoon between 1:30 and 6:00 PM.

Behavioral Analysis

The circular swimming pool (70 cm in diameter and 30 cm in height) was made of ivory-colored PVC, filled with water (25 ± 0.5 °C) made opaque with the Opacifier 631 to 12 cm below the edge of the wall. A circular goal platform (5 cm in diameter) laid 0.5 cm under the surface of the water and 7 cm from the wall. The device was placed in a regular room. Dropped into the water from a different quadrant on each trial, mice had to learn to navigate to the invisible platform using the spatial cues available in the room. After a three trial pretraining session to find out the procedural components of the task, the mice were given three consecutive trials a day for 4 days, according to the procedure described by Chapillon and Roullet (1996). After the third trial of the last session, the mice were submitted to a probe test for spatial bias. The platform was removed and the mouse, starting from the opposite quadrant, was allowed a 1-min search for the platform. The path was recorded on videotape and a spatial bias index was computed as the difference between the number of times an 8-cm-diameter annulus surrounding the former location of the platform was crossed and the mean number of crossings of three annuli, symmetrically laid out in the quadrants where the platform had never been, divided by the total number of annulus crossings.

Surgery, Drug Administration, and Histology

Selective and reversible inactivation of mossy fibers was obtained through direct infusion of DDC in the dorsal hippocampus. Under the same conditions, control mice received an infusion of Ca-EDTA. Mice were operated under deep chloral hydrate anesthesia (500 mg/kg). A holder made of methacrylate resin with two guide tubes spaced 4 mm apart and protruding 2 mm out of the base was fastened to the skull. The guide tubes were positioned according to stereotaxic coordinates from the atlas of Slotnick and Leonard (1975) (AP: 1.6 mm posterior to bregma; Lat: 2 mm; Vert: 1.6 below dura) so that the tip of the guide tubes was close to the dorsal part of the hippocampus, near the CA1 field. The mice were given a 5- to 7-day recovery period after surgery.

Just before the injections, two beveled injection tubes were introduced into the guides and their lengths adjusted so that the tip of the injection tube reached the CA3 pyramidal

layer. DDC and EDTA were administered as aqueous solutions (200 mM). Injections were monitored by a Bioblock infusion pump which infused 0.25 μl in 2 min simultaneously in both hippocampi. The injection tubes were maintained in place for 2 min after the end of the injection, so that the solution could not escape through the guide tube. Adjusted lengths of steel entomology pins smeared with paraffin oil were placed in the tubes to seal them between injections. After the injection, mice were replaced in their home cages for a 15-min period before testing.

Three days after the end of the experiments, mice were given a new injection of DDC; then, after a lapse of 20 min, they were perfused intracardially under lethal anesthesia with (a) 0.9% NaCl, (b) 0.1% sodium sulfide in phosphate buffer, (c) 3% glutaraldehyde fixative, and (d) 0.1% sodium sulfide. Their brains were then processed for Timm staining of the mossy fibers (Danscher & Zimmer, 1978) and counterstained with thionin for cellular bodies. The position of the cannulae was verified on 25-μm coronal sections. In all cases, damage to the dorsal cortex, close to the CA1 field, indicated that the guide tubes had been set at the right locations. Figure 2 presents the location of the tips of 10 guide tubes in each hemisphere that cover the entire area where injections were made. Bleaching of Timm stain of the mossy fibers (Frederickson et al., 1990) covered a large but not complete part of the mossy fiber synaptic field in the CA3 region. This suggests that the observed behavioral effects of reversible lesions of mossy fibers by the DDC result from moderate size focal lesions. Bleaching was not apparent in three mice. This could be due to an infusion problem liable to occur when many injections have already been made. Anyway, these animals did not appear to perform as outliers in their proper experimental group and consequently, it was more conservative not to discard them from the analysis.

Statistics

Box plots were used to look for outliers. A repeated measures ANOVA design was performed to analyze the effects of the treatment, of the session, and of their interaction on swimming latencies with the Multiple General Linear Model (Wilkinson, 1987). The influence of the treatments on the spatial bias index was analyzed by the nonparametric Mann–Whitney U test.

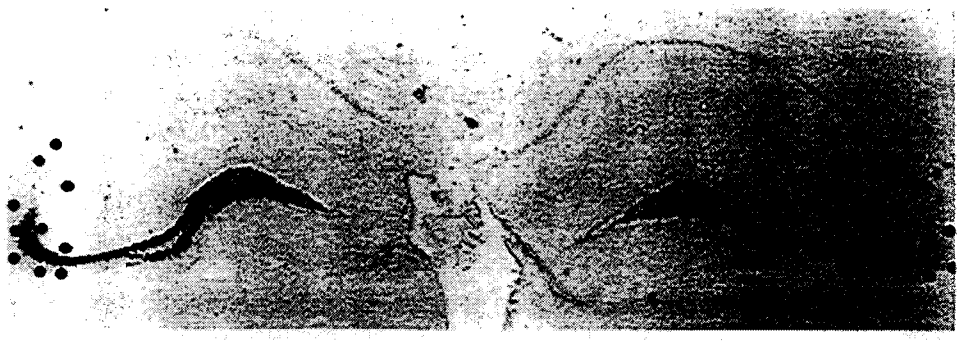

FIG. 2. Digitalized image of a coronal brain section showing (i) Timm bleaching of mossy fibers after DDC infusion on the right side compared to normal Timm staining on the left side and (ii) histological verification of cannulae placement in the dorsal hippocampus. Black dots indicate the location of 10 pairs of cannulae that cover the entire region where injections were made.

EXPERIMENT 1

Mice were first given a pretraining session with a visible platform (without drug injection) and then two series of four daily sessions with the submerged platform with a 72-h rest in between. The session started 15 min after the end of the injection. After the last trial of the last session, on day 4 of the first and second week, mice were submitted to the spatial probe test.

Thirty male mice were randomly assigned to three different groups: two groups of 12 mice each were equipped with guide tubes. These mice served alternatively as experimental and injected control groups according to a cross procedure. During the first week, group 1 mice were infused with EDTA. They were able to learn the task normally and were used as a control group to test the effects on learning of the DDC injected into mice of group 2. During the second week, group 2 mice received in turn the infusion of EDTA and served as controls of group 1 mice, which were then infused with DDC. The study of the performance of group 1 mice during the second week allowed analysis of the effects of DDC on long-term memory retrieval, whereas the study of the performance of group 2 mice checked for the reversibility of the inactivation of mossy fibers by the DDC and for the absence of residual effects according to the following predictions: (i) if mossy fiber synapses are not necessary to memory recall, group 1 mice should show normal performances under DDC during the second week of training; (ii) if DDC effects during week 1 are fully reversible, group 2 animals should show normal learning during week 2. One group of 6 nonimplanted and noninfused mice served as controls for the side effects of surgery and injections during the first week.

The escape performance of a session was the sum of the latencies of the three trials for a mouse within that session. Escape latencies were submitted to a $\log_{10}$ scale change in order to normalize the shape of the distributions and to homogenize the variances of the different experimental groups.

Results

Figure 3A shows that during the first week of training, mice displayed a significant linear improvement of their escape performance across sessions [$F(3,84) = 3.55$, $p = .018$] without significant treatment $\times$ session interaction [$F(6,84) = 0.783$, NS].

The ANOVA showed that there was no overall significant treatment effect either [$F(2,28) = 1.817$, NS]. Although it fluctuated greatly from the first to the second session, the performance of EDTA mice did not differ from that of controls [$F(1,16) = 0.005$, NS], which indicates that neither the implantation of the cannulae nor the infusion of EDTA significantly modified their learning performance. Mice infused with DDC during the first week of the experiment improved their performance more slowly than EDTA mice, but reached the same level of performance on the fourth day of training. Over the four sessions, the effect was nonsignificant [$F(1,23) = 3.324$, $p = .081$] but marginal. The probe test (Fig. 3B) demonstrated that the spatial performance of mice infused with EDTA during the first week of training did not differ from that of control mice [Mann–Whitney $U = 29.5$, $\chi^2 = 0.124$, $df = 1$, NS] and that they searched the platform at the right place. On the other hand, mice infused with DDC during the first week do not learn the location of the platform and searched for it everywhere in the water maze (DDC vs. EDTA: $U = 130.5$, $\chi^2 = 11.715$, $df = 1$, $p = .001$).

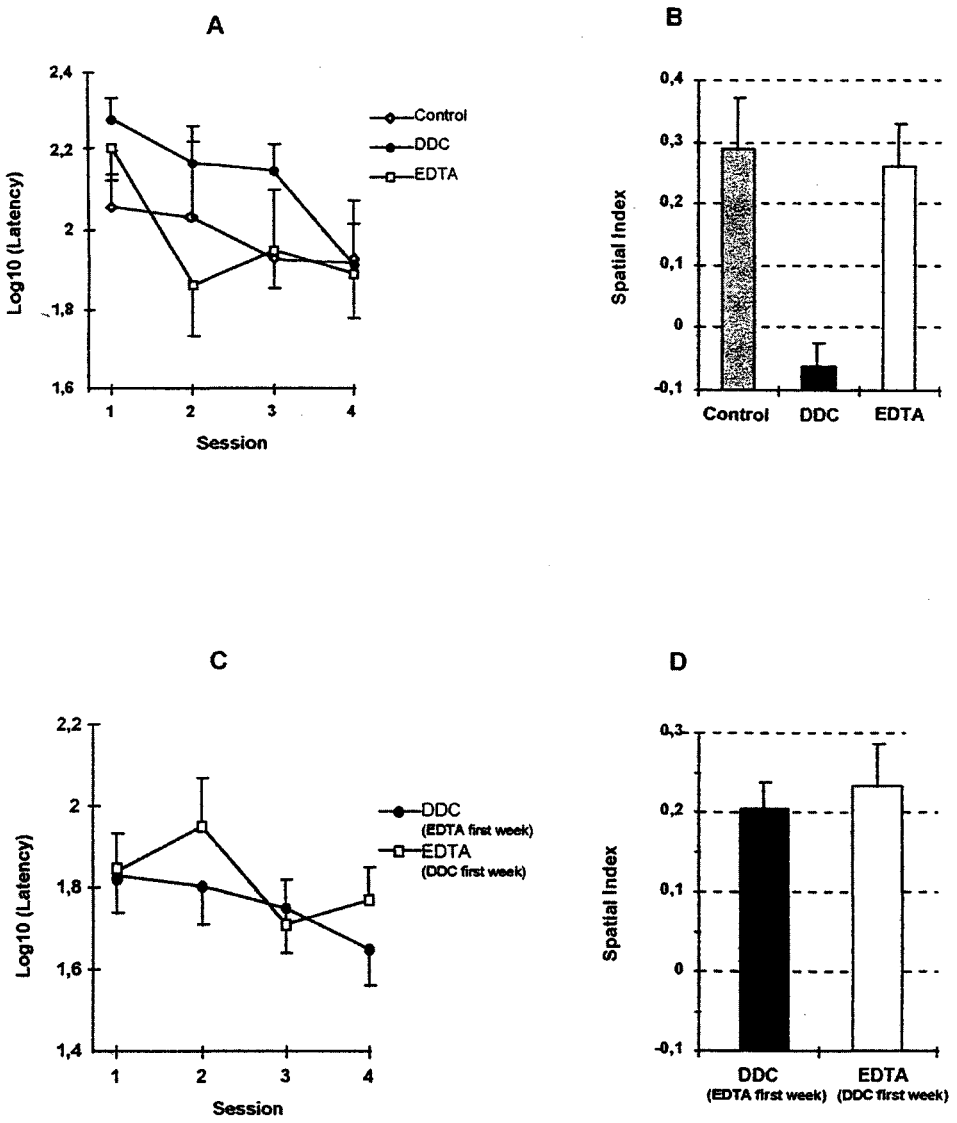

FIG. 3. (A) First week training. Log_{10} escape latencies (mean ± SEM) to find the platform in the Morris navigation task across sessions. Each latency is the sum of the three trial latencies within a session. (B) Spatial index values (mean ± SEM) during the probe test in the Morris navigation task at the end of the first week for control, EDTA, and DDC mice. (C) Second week training. Log_{10} escape latencies (mean ± SEM) to find the platform in the Morris navigation task after treatments have been crossed. (D) Spatial index values (mean ± SEM) during the probe test in the Morris navigation task at the end of the second week of training.

During the second week of the experiment (Fig. 3C), when treatments were rotated, mice continued to improve their escape latencies [$F(3,66) = 3.008, p = .036$] but, although latencies in EDTA mice fluctuated more than those of mice infused with DDC, both groups improved their performance at the same speed [F (1,22) = 0.315, NS]. Over the 2 weeks, mice infused with EDTA appeared, unexpectedly, more variable than their DDC counterparts.

The spatial probe test (Fig. 3D) showed that, whereas they were unable to learn the

spatial component of the task when they were infused with DDC during the first week of training, these mice can learn normally under EDTA during the second week. On the other hand, those which learned the spatial task under EDTA during the first week nevertheless displayed normal performances under DDC during the second week.

Discussion

The comparison of escape latencies and spatial index scores of EDTA and nonoperated, noninjected control mice proves that surgery and injection have no side effects and that EDTA has no effect on mossy fiber synapses. Consequently, EDTA mice are the right control for DDC mice. These results establish also that the effect of DDC is fully reversible, without any aftereffects, since DDC mice which were unable to learn the location of the platform during the first week of training learned normally during the next week under EDTA.

The analysis of escape latencies shows that the inactivation of the mossy fibers by DDC might slow the improvement of latencies, at least during the previous stages, but was unable to prevent it. Nevertheless the improvement of escape performance across sessions does not imply that mice learned to locate the platform. They might also have learned only that somewhere under the water, at a certain distance from the wall, there is a platform they can climb for a rest, so that they searched actively and efficiently for it, without being able to navigate there directly.

Actually, analysis of the spatial index shows that the DDC, by selectively blocking the synapses of the mossy fibers on the CA3 neurons, prevents learning the spatial component of the task, i.e., learning the location of the platform, whereas it does not impede learning the sensorimotor and procedural components of the task (hunting actively everywhere for a platform), which results in the improvement of escape latencies over sessions. These results support the involvement of the hippocampal structure, specially the CA3 region, in the specific learning of the spatial component of the task. This corresponds to a first dissociation.

Inactivation of mossy fibers by the DDC also reveals a second dissociation. According to the model presented by Treves and Rolls (1992, 1994), these results show that, whereas the activity of the mossy fibers is essential to the learning process, it is not necessarily involved in memory recall, since blocking mossy fiber activity during the second week of training does not prevent the recall of spatial information stored previously during the first week. Nevertheless, this last point holds only if it is considered that the information is still stored in the modifiable synapses about 72 h after the learning session. An alternative hypothesis to account for these results could be that instead of preventing the acquisition of spatial information, mossy fiber transient inactivation could interfere with the early processes of memory consolidation and thus impair memory storage. Posttrial DDC injections, which leave mossy fibers functional during acquisition but block their activity during the first 45 min of memory consolidation, would help answer this question.

EXPERIMENT 2

This second experiment aimed at dissociating the effects of mossy fiber synapse inactivation by DDC on the learning, memory consolidation, and recall processes.

Thirty-two mice, equipped with injection tubes, were allotted to four groups receiving DDC or EDTA infusions either 15 min before the first trial of a session, in order to be active during the learning phase, or immediately after the last trial, to affect the initial phase of the memory consolidation process.

At the end of the learning session on the fourth day, after a 15-min pause, the mice were given a spatial probe test. The two postsession injection groups received a last DDC or EDTA infusion immediately after the last learning trial so that the effect of mossy fibers inactivation on memory recall could be checked 15 min later during the probe test. During this second experiment, the water temperature was lowered from 25 to 23°C in order to improve the motivation of animals to escape.

Results

Figure 4A showed that presession injected mice presented a significant global improvement of their latencies to find the platform from the first to the fourth sessions whatever treatment they received [$F(3,33) = 9.477$, $p < .001$]. Comparison of the performances between groups which received either DDC or EDTA before the training sessions supported the results of the first experiment; DDC mice showed longer latencies than EDTA mice before finding the platform, although the difference was again nonsignificant [$F(1,11) = 3.697$, $p = .081$] but marginal. As previously, they reached similar performance on the fourth session [$F(1,11) = 0.578$, NS]. Results of the spatial probe test (Fig. 4B) showed that mice which received a DDC infusion before the learning sessions did not learn the spatial location of the platform, whereas in the same conditions, EDTA mice could learn [$U = 41.5$, $\chi^2 = 8.6$, $df = 1$, $p = .003$].

On the other hand, when the infusions were given immediately after each learning session, DDC and EDTA did not affect consolidation in a different manner. The latencies of mice which received the DDC (Fig. 4C) did not differ globally from those which received EDTA [$F(1,14) = 0.619$, NS] and both improved significantly their performance across sessions [$F(3,42) = 8.887$, $p < .001$]. The difference observed on the first trial between the posttrial DDC and EDTA mice cannot be attributed to the effect of the molecules, since they had not received any infusion at that time. The results of the spatial probe test (Fig. 4D) showed that the spatial index of DDC mice did not differ from that of EDTA controls [$U = 34.5$, $\chi^2 = 0.069$, $df = 1$, NS].

Discussion

These results show that the inactivation of mossy fibers during the early stages of memory consolidation disrupted neither the storage of information acquired during the session immediately prior to the injection nor the recall of this information. They also confirm the findings of the first experiment; whereas mossy fiber inactivation during learning impaired the initial performance but did not prevent mice from improving their escape latencies during the next sessions, it nevertheless made spatial learning impossible.

GENERAL DISCUSSION

The results of these two experiments first confirm already known phenomena: (i) they corroborate the involvement of the hippocampus and particularly of the CA3 region in

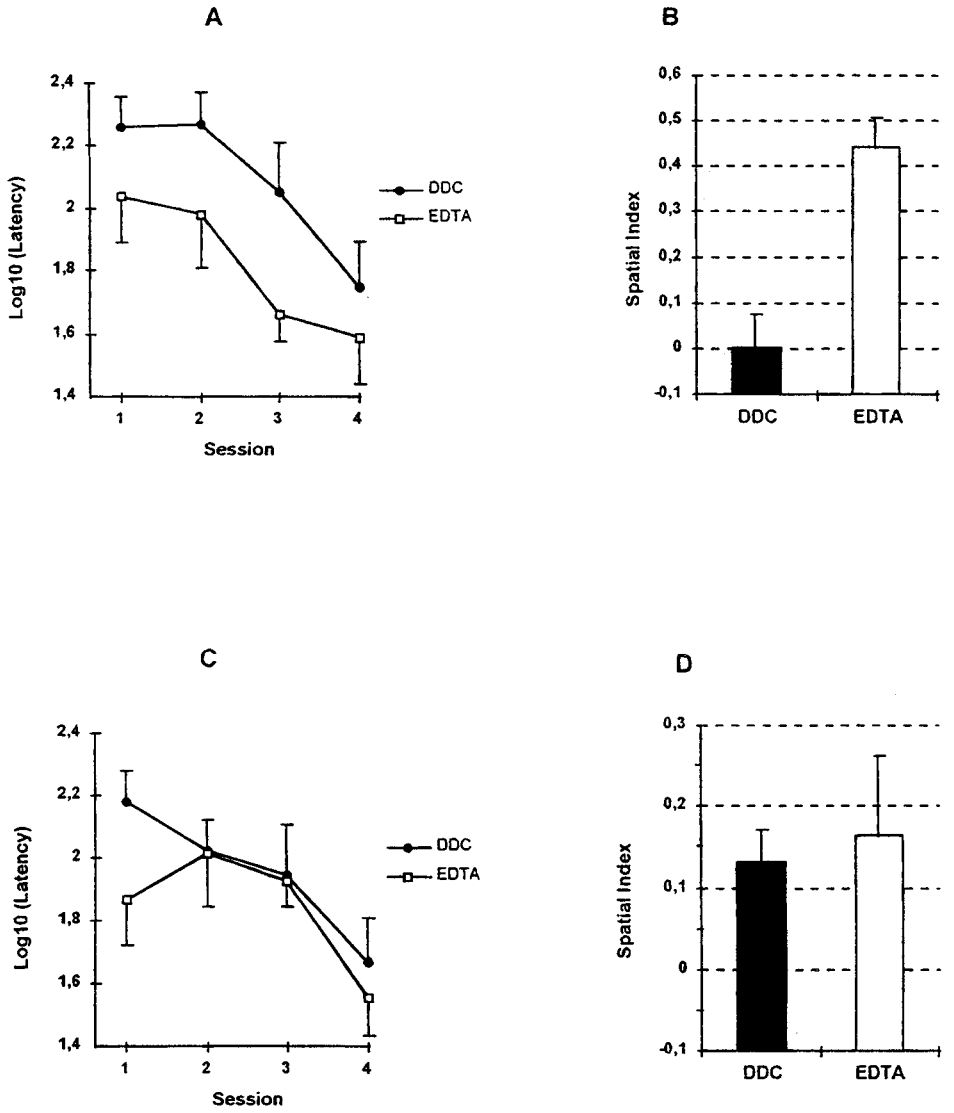

FIG. 4. (A) Presession injection. Log_{10} escape latencies (mean ± SEM) to find the platform in the Morris navigation task. (B) Spatial index values (mean ± SEM) during the probe test in the Morris navigation task. (C) Postsession injection. Log_{10} escape latencies (mean ± SEM) to find the platform in the Morris navigation task. (D) Spatial index values (mean ± SEM) during the probe test in the Morris navigation task.

the learning of a spatial location in the Morris navigation task, (ii) they validate the dissociation between procedural and sensorimotor learning on one side and spatial learning on the other side, as far as hippocampal functions are concerned. Above all, they bring new findings of potential importance to understanding the role played by mossy fibers in the learning and memory processes of a spatial task.

Since the first report on cognitive deficits due to bilateral hippocampal lesions by Brenda Milner in the fifties (Scoville & Milner, 1957), neuropsychologists have clearly demonstrated that whereas hippocampal subjects suffer anterograde amnesia and spatio-temporal desorientation, they are still capable of forms of procedural learning which

concern the acquisition rules or the implementation of sensorimotor abilities. On these lines, it is more and more widely acknowledged by specialists who study animal behavior that the escape latency in the spatial version of the Morris task is a rather complex behavioral performance involving various processes (attention, motivation, rule understanding, and specific and strain-specific strategies) so that it cannot represent a reliable assessment of the processes that control spatial orientation. Only the spatial probe test will give an appropriate measurement of the spatial learning ability of the subject [see Upchurch and Wehner (1989) for an earlier discussion of this point and Lipp and Wolfer (1998) for a more recent one]. Whishaw (1989) brought forward experimental arguments concerning the need for dissociating between performance and learning deficits in spatial navigation tasks in rats submitted to cholinergic blockade. Numerous experiments show that brain-lesioned rats (hippocampus, neocortex, subiculum, caudate putamen) can significantly improve their escape latencies toward an invisible platform in the Morris swimming pool, although they do not reach the same level of performance as controls. On the other hand, they display a significant deficit in the probe test (see for example Morris, 1990; Moser et al., 1993; Whishaw et al., 1987). Similarly, Frederickson et al. (1990) noticed that whereas the rat's ability to return directly to the specific platform position on the second trial of the delayed matching to sample task in the Morris swimming pool was severely impaired by hippocampal infusions of DDC, recall of the general procedure for solving the task, however, was relatively unaffected by the drug. Namely, no trained rat ever reverted to the naïve strategy of swimming around the tank in search of escape, for the duration of the trial. Nevertheless, it has been shown that overtraining hippocampal rats allows them to improve even more their escape latencies and finally their spatial performance in the probe test (Morris, 1990). In this experiment however, mice were given moderate training so that their performances ranged from a mean value of about 67 s on the first session to 15 s in the last session, which enabled few mice to have direct paths toward the platform. In this respect, our results match those of Mizumori, Perez, Alvaro, Barnes, and McNaughton (1990), who showed that in rats, reversible inactivation of the medial septum by tetracaïne impairs spatial learning on the radial maze, whereas it produces only a significant retardation of learning followed by a clear improvement over trials in the spatial reference memory task with the same experimental paradigm.

Although the effects of DDC observed in the present study are interpreted as a consequence of the binding of the zinc in hippocampal mossy fiber synapses, other possible effects of DDC have been investigated. DDC can also act as a dopamine-β-hydroxylase inhibitor which has been shown to reduce whole brain norepinephrine *in vivo* (Haycock et al., 1978; Frieder & Allweis, 1982). The effects of systemic injections of DDC have been investigated in one-trial avoidance tasks in rats. It has been shown that DDC injected prior to training does not impair learning and short-term memory performance (Hamburg & Cohen, 1973; Stein et al., 1975; Solanto & Hamburg, 1979; Frieder & Allweis, 1982). The effects of DDC observed in our experiments or those of Frederickson et al. (1990), however, cannot be explained by their effects through the noradrenergic system for various reasons. First, as underlined by Haycock et al. (1978), the lack of effect of other dopamine-β-hydroxylase inhibitors makes it difficult to attribute the amnestic effects of DDC solely to catecholaminergic effects. Second, the absence of effect of systemic or intracisternal injections of DDC on short-term memory, whereas they impair memory consolidation or recall processes, indicates clearly that the target is not the hippocampus. These effects

are in the opposite direction of those resulting from hippocampal insults. Third, the timing of the pharmacological interventions and their effects in these experiments are very different from those observed in our experiments. The delay (hours to days) between the injections and their effects as well as their duration are clearly incompatible with those of intrahippocampal infusions. Fourth, the cognitive processes involved in these experiments and in our learning task are also different and there is no evidence that the hippocampus is necessary to acquire a one-trial associative conditioning when there is no delay between the occurrence of the behavior to be suppressed (step down or step through) or the place to be passively avoided and the administration of the punishment. All these arguments and others developed by Danscher et al. (1973), namely, the fact that other zinc chelating agents, dithizone (Fleischhauer & Ohnesorge, 1958) and oxine (Danscher & Fredens, 1972), have the same effect on the Timm stain and have somewhat similar behavioral effects supports our claim that the effects of intrahippocampal infusions of DDC on spatial learning are mediated through the binding of zinc, which interferes with transmission in the mossy fiber synapses of the CA3–CA4 region.

On the whole, our results demonstrate that reversible inactivation of the mossy fibers by DDC disrupts the spatial learning process, whereas it has no effect on memory consolidation or memory recall in either working or reference memory. This new dissociation supports the selective involvement of the mossy fiber synapses in the learning process of a spatial location, whereas they are not necessary to memory recall, in accordance with the model developed by Rolls (1994) and Treves and Rolls (1992, 1994), in which the CA3 region of the hippocampus is supposed to act as an autoassociation memory matrix. The consequences of the inactivation of mossy fibers can be paralleled with those of reversible inactivation of the medial septal area (MSA), which sends cholinergic inputs either directly to the hippocampus via the fimbria fornix or indirectly through layer II of the entorhinal cortex. Our results are again consistent with those of Mizumori et al. (1990), since inactivation of MSA before learning increased error numbers during the test, whereas inactivation of the MSA after the initial learning phase had no effect on the test and thus can be said not to affect consolidation. On the other hand, they differ from those of Rashidy-Pour, Motghed-Larijani, and Bures (1996), which show that reversible inactivation of the MSA by tetrodotoxin impairs consolidation of a passive avoidance learning task in rats when administered 5 to 90 min after a single acquisition trial. Such inconsistencies stress either possible differences in the effects of the chemicals used in these studies or the diversity of the neural mechanisms underlying the two learning tasks, rather than a weakness of the reversible inactivation methodology, which proved extremely selective and efficient.

Genetic studies by Heimrich et al. (1985), Crusio et al. (1986), and Lassalle et al. (1999) indicate that size variations of the various hippocampal mossy fiber layers (suprapyramidal, intra-infrapyramidal, and CA4) are based on different genetic architectures. As already underlined, they also show genetic correlations with various behavioral processes, the physiological and cognitive bases of which are poorly understood. In most cases, behavioral variation correlates with the size of the intra-infrapyramidal mossy fiber projection rather than with the entire mossy fiber area. Therefore, it is important to know more about details of the role played by hippocampal mossy fibers in hippocampal functionning to decipher these correlations. As our results show that mossy fibers are involved in the storage of episodic memories, the focus should then be put on physiological mechanisms

that could explain that mice with a large intra-infrapyramidal projection are better performers. For instance, it would be relevant to know if mossy fibers that synapse on the basal dendrites of pyramidal cells are more efficient in teaching these cells or inducing mossy fiber potentiation, as has been shown in the CA1 field by Capocchi et al. (1992), Karbara and Leung (1993), and Arai et al. (1994). Also, as CA3, like CA1, pyramids are complex spike cells that have been considered basic elements of the spatial neural representations, research should be undertaken to discover whether they are involved only for the tuning of these place cells or whether they might as well be involved in nonspatial learning processes. Further studies, investigating the effects of reversible lesions in tasks involving spatial and nonspatial components, are currently under way in our laboratory that should help clarify this point.

REFERENCES

Arai, A., Black, J., & Lynch, G. (1994). Origins of the variations in long-term potentiation between synapses in the basal versus apical dendrites of hippocampal neurons. *Hippocampus*, **4,** 1–10.

Blair, H. T., & Sharp, P. E. (1996). Visual and vestibular influences on head-direction cells in the anterior thalamus of the rat. *Behavioral Neuroscience*, **110,** 643–660.

Bures, J., & Buresova, O. (1990). Reversible lesions allow reinterpretation of system level studies of brain mechanisms of behavior. *Concepts in Neurosciences*, **1,** 69–89.

Burgess, N., & O'Keefe, J. (1996). Neuronal computations underlying the firing of place cells and their role in navigation. *Hippocampus*, **7,** 749–762.

Capocchi, G., Zampolini, M., & Larson, J. (1992). Theta burst stimulation is optimal for induction of LTP at both apical and basal dendritic synapses on hippocampal CA1 neurons. *Brain Research*, **591,** 332–336.

Chapillon, P., & Roullet, P. (1996). Ontogeny of orientation and spatial learning on the radial maze in mice. *Developmental Psychobiology*, **28,** 429–442.

Cressant, A., Muller, R. U., & Poucet, B. (1997). Failure of centrally placed objects to control the firing fields of hippocampal place cells. *The Journal of Neuroscience*, **17,** 2531–2542.

Crusio, W. E., Genthner-Grimm, G., & Schwegler, H. (1986). A quantitative-genetic analysis of hippocampal variation in the mouse, *Journal of Neurogenetics*, **3,** 203–214.

Crusio, W. E., Schwegler, H., & Van Abeelen, J. H. F. (1989). Behavioral responses to novelty and structural variation of the hippocampus in mice. II. Multivariate genetic analysis. *Behavioural Brain Research*, **32,** 81–88.

Danscher, G., & Fredens, K. (1972). The effect of oxine and alloxan on the sulfide silver stainability of the rat brain. *Histochemie*, **30,** 307–314.

Danscher, G., Haug, F.-M. S., & Fredens K. (1973). Effect of diethyldithiocarbamate (DEDTC) on sulphide silver stained boutons. Reversible blocking of Timm's sulfide silver stain for "heavy" metals in DEDTC treated rats (light microscopy). *Experimental Brain Research*, **16,** 521–532.

Danscher, G., & Zimmer, J. (1978). An improved Timm sulphide silver method for light and electron microscopic localization of heavy metals in biological tissues. *Brain Research*, **425,** 27–40.

Dudchenko, P. A., & Taube, J. S. (1997). Correlation between head direction cell activity and spatial behavior on a radial arm maze. *Behavioral Neuroscience*, **111,** 3–19.

Fleischhauer, K., & Ohnesorge, F. K. (1958). Zur Pharmakologie des dithizon. *Naunyn-Schmiedeberg's Archiv des Experimentalen und Pathologisches Pharmakologie*, **235,** 63–77.

Fredens, K., & Dansher, G. (1973). The effect of intravital chelation with dimercaprol, calcium disodium edetate, 1-10-phenantroline and 2,2′-dipyridyl on the sulfide silver stainability of the rat brain. *Histochemie*, **37,** 321–331.

Frederickson, R. E., Frederickson, C. J., & Dansher, G. (1990). In situ binding of bouton zinc reversibility disrupts performance on a spatial memory task. *Behavioural Brain Research*, **38,** 25–33.

Frieder, B., & Allweis, C. (1982). Memory consolidation: further evidence for the four-phase model from the time-courses of diethyldithiocarbamate and ethacrinic acid amnesias. *Physiology and Behavior*, **29,** 1071–1075.

Gallo, M., & Candida, A. (1995). Reversible inactivation of dorsal hippocampus by tetrodotoxin impairs blocking of taste aversion selectively during the acqisition but not the retrieval of rats. *Neuroscience Letters*, **186,** 1–4.

Guillot, P. V., Roubertoux, P. L., & Crusio, W. E. (1994). Hippocampal mossy fiber distributions and intermale aggression in seven inbred mouse strains. *Brain Research*, **660,** 167–169.

Hamburg, M. D., & Cohen, R. P. (1973). Memory access pathway: role of adrenergic versus cholinergic neurons. *Pharmacology, Biochemistry and Behavior*, **1,** 295–300.

Haug, F.-M. S. (1967). Electron microscopal localization of the zinc in hippocampal mossy fibre synapses by a modified sulfide silver procedure. *Histochemie*, **8,** 355–368.

Hausheer-Zarmakupi, Z., Wolfer, D. P., Leisinger-Trigona, M. C., & Lipp, H. P. (1996). Selective breeding for extremes in open-field activity of mice entails a differenciation of hippocampal mossy fibers. *Behavioral Genetics*, **26,** 167–176.

Haycock, J. W., Van Buskirk, R., Gold, P. E., & McGaugh, J. L. (1978). Effects of diethyldithiocarbamate and fusaric acid upon memory storage processes in rats. *European Journal of Pharmacology*, **51,** 261–273.

Heimrich, B., Schwegler, H., & Crusio, W. E. (1985). Hippocampal variation between the inbred mouse strains C3H/HeJ and DBA/2: A quantitative-genetic analysis. *Journal of Neurogenetics*, **2,** 389–401

Kaibara, T., & Leung, L. S. (1993). Basal versus apical dendritic long-term potentiation of commissural afferents to hippocampal CA1: A current-source density study. *Journal of Neurosciences*, **13,** 2391–2404.

Lassalle, J. M. (1996). Neurogenetic bases of cognition: Facts and hypotheses. *Behavioural Processes*, **35,** 5–18.

Lassalle, J. M., Halley, H., Milhaud, J. M., & Roullet, P. (1999). Genetic architecture of the hippocampal mossy fiber subfields in the BXD RI mouse strain series: A preliminary QTL analysis. *Behavior Genetics*, **29,** 273–282.

Lipp, H. P., Schwegler, H., Crusio, W. E., Wolfer, D. P., Leisinger-Trigona, M.-C., Heimrich, B., & Driscoll, P. (1989). Using genetically-defined rodent strains for the identification of hippocampal traits relevant for two-way avoidance behavior: a non-invasive approach. *Experientia*, **45,** 845–859.

Lipp, H. P., & Wolfer, D. P. (1998). Genetically modified mice and cognition. *Current Opinion in Neurobiology*, **8,** 272–280.

Mizumori, S. J. Y., Perez, G. M., Alvaro, M. C., Barnes, C. A., & Mc Naughton, B. L. (1990). Reversible inactivation of the medial septum differentially affects two forms of learning in rats. *Brain Research*, **528,** 12–20.

Morris, R. G. M. (1990). Toward a representational hypothesis of the role of hippocampal synaptic plasticity in spatial and other forms of learning. *Cold Spring Harbor Symposia on Quantitative Biology*, **55,** 161–173.

Morris, R. G. M., Garrud, P., Rawlins, J. N. P., & O'Keefe, J. (1992). Place navigation in rats with hippocampal lesions. *Nature*, **297,** 681–683.

Moser, E., Moser, M.-B., & Andersen, P. (1993). Spatial learning impairment parallels the magnitude of dorsal hippocampal lesions, but is hardly present following ventral lesions. *The Journal of Neurosciences*, **13,** 3916–3925.

McNaughton, B. L., & Smolensky, P. (1991). Connectionist and neural modeling: Converging in the hippocampus. *In* R. G. Liister & H. J. Weingartner (Eds.), *Perspectives on cognitive neurosciences* (pp. 93–110). New York: Oxford Univ. Press.

O'Keefe, J. (1991). The hippocampal cognitive map and navigational strategies. *In* J. Paillard (Ed.), *Brain and space*. Oxford: Oxford Univ. Press.

O'Keefe, J., & Burgess, N. (1996). Geometric determinants of the place fields of hippocampal neurons. *Nature*, **381,** 425–428.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, **34,** 171–175.

Perez-Clausel, J., & Dansher, G. (1985). Intravesicular localization of zinc in rat telencephalic boutons. A histochemical study. *Brain Research*, **337,** 91–98.

Rashidy-Pour, A., Motghed-Larijani, Z., & Bures, J. (1996). Reversible inactivation of the medial septal area impairs consolidation but not retrieval of passive avoidance learning in rats. *Behavioural Brain Research*, **72,** 185–188.

Rolls, E. T. (1994). Functions of the primate hippocampus in spatial processing and memory. *In* J. Paillard (Ed.), *Brain and space* (pp. 353–376) Oxford: Oxford Univ. Press.

Roullet, P., & Lassalle, J. M. (1990). Genetic variation, hippocampal mossy fibres distribution, novelty reactions and spatial representation in mice. *Behavioural Brain Research*, **48,** 77–85.

Schwegler, H., & Crusio, W. E. (1995). Correlations between radial-maze learning and structural variations of septum and hippocampus in rodents. *Behavioural Brain Research*, **67,** 29–41.

Schwegler, H., Crusio, W. E., Lipp, H. P., Brust, I., & Mueller, G. G. (1991). Early postnatal hyperthyroidism alters hippocampal circuitry and improves radial maze learning in adult mice. *Journal of Neurosciences*, **11,** 2102–2106.

Scoville, W. B., & Milner, B., (1957). Loss of recent memory after bilateral hippocampal lesion. *Journal of Neurology, Neurosurgery and Psychiatry*, **20,** 11–21.

Slotnik, B. M., & Leonard, C. M. (1975). *A stereotaxic atlas of the albino mouse forebrain.* Rockville, MD: U.S. Department of Health, Education and Welfare.

Solanto, M. V., & Hamburg, M. D. (1979). DDC-induced amnesia and norepinephrine: A correlated behavioral-biochemical analysis. *Psychopharmacology*, **66,** 167–170.

Stein, L., Belluzzi, J. D., & Wise, C. D. (1975). Memory enhancement by central administration of norepinephrine. *Brain Research*, **84,** 329–335.

Sutherland, R. J., & Rudy, J. W. (1988). Place learning in the Morris place navigation task is impaired by damage to the hippocampal formation even if the temporal demands are reduced. *Psychobiology*, **16,** 157–163.

Taube, J. S., Muller, R. U., & Rank, J. B., Jr. (1990) Head direction cells recorded from the postsubiculum in freely moving rats: I. Description and quantitative analysis. *Journal of Neuroscience*, **10,** 420–435.

Treves, A., & Rolls, E. T. (1992). Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus*, **2,** 189–199.

Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, **4,** 374–391.

Upchurch, M., & Wehner, J. (1989). Inheritance of spatial learning ability in inbred mice: a classical genetic analysis. *Behavioral Neuroscience*, **103,** 1251–1258.

Whishaw, I. Q. (1989). Dissociating performance and learning deficits on spatial navigation tasks in rats subjected to cholinergic muscarinic blockade. *Brain Research Bulletin*, **23,** 347–358.

Whishaw, I. Q., Mittleman, G., Bunch, S. T., & Dunnett, S. B. (1987). Impairments in the acquisition, retention and selection of spatial navigation strategies after medial caudate-putamen lesions in rats. *Behavioural Brain Research*, **24,** 125–138.

Wilkinson, L. (1987). *SYSTAT: The System for Statistics.* Evanston: II: SYSTAT Inc.

Wimer, R. E., & Wimer C. C. (1985). Animal behavior genetics: A search for the biological foundations of behavior. *Annual Review of Psychology*, **36,** 171–218.

Computational constraints that may have favoured the emergence of neocortical lamination

Alessandro Treves
SISSA - Programme in Neuroscience, via Beirut 4, 34014 Trieste, Italy, ale@sissa.it

Both of the major qualitative changes in the cerebral cortex, at the transition from early reptilian ancestors to primordial mammals, involve the insertion of a layer of granule cells. In the medial cortex, the detachment of the dentate gyrus defines the organization of the modern mammalian hippocampus, a new organization which may be understood as affording a quantitative computational advantage.

Concurrently, the dorsal cortex granulates, that is the principal layer of pyramidal cells is split by the insertion of a new layer of excitatory, but intrinsic, granule cells. An hypothesis is formulated that explains the emergence of isocortical lamination in the evolution of mammals as necessary to support fine topography in their sensory maps. Fine topography implies a generic distinction between "where" information, explicitly mapped on the cortical sheet, and "what" information, represented in a distributed fashion as a distinct firing pattern across neurons. These patterns can be stored on recurrent collaterals in the cortex, and such memory can help substantially in the analysis of current sensory input. The analysis of suitable network models, and their simulation, demonstrates that a nonlaminated patch of cortex must compromise between transmitting "where" information or retrieving "what" information; whereas the differentiation of a granular layer affords, again, a quantitative advantage. The further connectivity differentiation between infragranular and supragranular pyramidal layers is shown to match the mix of "what" and "where" information optimal for their respective target structures.

1. Introduction

The evolution of mammalian cortex. Mammals originate from the therapsids, one order among the first amniotes, or early reptiles, as they are commonly referred to. They are estimated to have radiated away from other early reptilian lineages, including the anapsids (the progenitors of modern turtles) and diapsids (out of which other modern reptilians, as well as birds, derive) some three hundred million years ago [1]. Perhaps mammals emerged as a fully differentiated class out of the third-to-last of the great extinctions, in the Triassic period. The changes in the organization of the nervous system, that mark the transition from proto-reptilian ancestors to early mammals, can be reconstructed only indirectly. Along with supporting arguments from the examination of endocasts (the inside of fossil skulls; [2]) and of presumed behavioural patterns [3], the main line of evidence is the comparative anatomy of present day species [4]. Among a variety of *quantitative* changes in the relative development of different structures, changes that have been extended, accelerated and diversified during the entire course of mammalian evolution [5], two major *qualitative* changes stand out in the forebrain, two new features that, once established, characterize the cortex of mammals as distinct from that of reptilians and birds. Both these changes involve the introduction of a new "input" layer of granule cells. In the first case, it is the medial pallium (the medial part of the upper surface of each cerebral emisphere, as it bulges out of the forebrain) that reorganizes into the modern-day mammalian hippocampus. The crucial step is the detachment of the most medial portion, that loses both its continuity with the rest of the cortex at the hippocampal sulcus, and its projections to dorso-lateral cortex [6]. The rest of the medial cortex becomes Ammon's horn, and retains the distinctly cortical pyramidal cells, while the detached cortex becomes the dentate gyrus, with its population of granule cells, that project now, as a sort of pre-processing stage, to the pyramidal cells of field CA3 [7]. In the second case, it is the dorsal pallium (the central part of the upper surface) that reorganizes internally, to become the cerebral neocortex. Aside from special cases, most mammalian neocortices display the characteristic isocortical pattern of lamination, or organization into distinct layers of cells (traditionally classsified as 6, in some cases with sublayers). The crucial step, here, appears to be the emergence, particularly evident in primary sensory cortices, of a layer of non-pyramidal cells (called spiny stellate cells, or granules) inserted in between the pyramidal cells of the infragranular and supragranular layers. This is layer IV, where the main ascending inputs to cortex terminate [8].

An information-theoretical advantage. What is the evolutionary advantage, for mammals, brought about by these changes?

In the case of the hippocampus, attempts to account for its remarkable internal organization have been based, since the seminal paper by David Marr [9], on the computational analysis of the role of the hippocampus in memory. The hippocampus is important for spatial memory also in birds. A reasonable hypothesis is that the "invention" of the dentate gyrus enhances its capability, in mammals, to serve as a memory store. Working with Edmund Rolls and building on the approach outlined by David Marr, I have proposed 10 years ago [10] that the new input to CA3 pyramidal cells from the mossy fibers (the axons of the dentate granule cells) serves to create memory representations in CA3 richer in information content than they could have been otherwise. The crucial prediction of this proposal was that the inactivation of the mossy fiber synapses should impair the formation of new hippocampal dependent memories, but *not* the retrieval of previously stored ones. This prediction has recently been verified [11] in mice. Thus a quantitative, information-theoretical advantage may have favored a qualitative change, such as the insertion of the dentate gyrus in the hippocampal circuitry. This idea, still to be tested further, raises the issue of whether also the insertion of layer IV in the isocortex might be accounted for in quantitative, information-theoretical terms.

Layers and maps. It has long been hypothesized that isocortical lamination appeared together with fine topography in cortical sensory maps [12], pointing at a close relationship between the two phenomena. All of the cortex, which develops from the upper half of the neural tube of the embryo, has been proposed to have been, originally, sensory, with the motor cortex differentiating from the somatosensory portion [13,14]. In early mammals, the main part of the cortex was devoted to the olfactory system, which is not topographic, and whose piriform cortex has never acquired isocortical lamination [15]. The rest of the cortex was largely allocated to the somatosensory, visual and auditory system, perhaps with just one topographic area, or map, each [4]. Each sensory map thus received its inputs directly from a corresponding portion of the thalamus, as opposed to the network of cortico-cortical connections which has been greatly expanded [16,17] by the evolution of multiple, hierarchically organized cortical areas in each sensory system [18,19]. In the thalamus, a distinction has been drawn [20] between its matrix and core nuclei. The matrix, the originally prevalent system, projects diffusely to the upper cortical layers; while the core nuclei, which specialize and become dominant in more advanced species [21], project with topographic precision to layer IV, although their axons contact, there, also the dendrites of pyramidal cells whose somata lie in the upper and deep layers.

2. A functional hypothesis.

The crucial aspect of fine topography in sensory cortices is the precise correspondence between the location of a cortical neuron and the location, on the array of sensory receptors, where a stimulus can best activate that neuron. Simple visual and somatosensory cortices thus comprise 2D maps of the retina and of the body surface, while auditory cortices map sound frequency in 1 dimension, and what is mapped in the other dimension is not quite clear [22]. Some of the parameters characterizing a stimulus, those reflected in the position of the receptors it activates, are therefore represented continuously on the cortical sheet. I will define them as providing *positional* information. Other parameters, which contribute to identify the stimulus, are not explicitly mapped on the cortex. For example, the exact nature of a tactile stimulus at a fixed spot on the skin, whether it is punctuate or transient or vibrating, and to what extent, are reflected in the exact pattern of activated receptors, and of activated neurons in the cortex, but not directly in the position on the cortical sheet. I will define these parameters as providing *identity* information. Advanced cortices, like the primary visual cortex of primates, include complications due to the attempt to map additional parameters on the sheet, like ocular dominance or orientation, in addition to position on the retina. This leads to the formation of so-called columns, or wrapped dimensions, and to the differentiation of layer IV in multiple sublayers. They should be regarded as specializations, which likely came much after the basic cortical lamination scheme had been laid out.

The sensory cortices of early mammals therefore received from the thalamus, and had to analyse, information about sensory stimuli of two basic kinds: positional or *where* information, I_p, and identity or *what* information, I_i. These two kinds differ also in the extent to which cortex can contribute to the analysis of the stimulus. Positional information is already represented explicitly on the receptor array, and then in the thalamus, and each relay stage can only degrade it. At best, the cortex can try to maintain the spatial resolution with which the position of a stimulus is specified by the activation of thalamic neurons: if these code it inaccurately, there is no way the cortex can reconstruct it any better, because any other position would be just as plausible. The identity of a stimulus, however, may be coded inaccurately by the thalamus, with considerable noise, occlusion and variability, and the cortex can reconstruct it from such partial information. This is made possible by the storage of previous sensory events in terms of distributed efficacy modifications in synaptic systems, in particular on the recurrent collaterals connecting pyramidal cells in sensory cortex. Neural network models of autoassociative memories [9,23] have demonstrated how simple "Hebbian" rules modelling associative synaptic plasticity can induce weight changes that lead to the formation of dynamical attractors [24]. Once an attractor has been formed, a partial cue corresponding e.g. to a noisy or occluded version of a stimulus can take the recurrent network within its basin of attraction, and hence lead to a pattern of activation of cortical neurons, which represents the stored identity of the original stimulus. Thus by exploiting dishomogeneities in the input statistics - some patterns of activity, those that have been stored, are more "plausible" than others - the cortex can reconstruct the identity of stimuli, over and beyond the partial information provided by the thalamus. This analysis of current sensory experience in the light of previous experience is hypothesized here to be the generic function of the cortex, which thus blends perception with memory [25]. Specialized to the olfactory sense, this function does not seem to require new cortical machinery to be carried out efficiently. I explore here the possibility that a novel circuitry is instead advantageous, when the generic function is specialized to topographic sensory systems, which have to relay both where and what information, I_p and I_i.

3. The simulated model of an isocortical patch.

Does preserving accurate coding of position conflict with the analysis of stimulus identity? This is obviously a quantitative question, which has to be addressed with a suitable neural network model. The following is an attempt to define a minimal model, which still includes in simplified form all the relevant ingredients.

A patch of cortex is modelled as a wafer of 3 arrays, each with $N \times N$ units. Each unit receives C_{ff} feedforward connections from a further array of $N \times N$ "thalamic" units, and C_{rc} recurrent connections from other units in the patch (Fig. 1). Both sets of connections are assigned to each receiving unit at random, with a Gaussian probability in register with the unit itself, and of

width S_{ff} and S_{rc}, respectively[1]. To model, initially, a *uniform*, non-laminated patch, the 3 arrays are identical in properties and connectivity, so the C_{rc} recurrent connections each unit receives are drawn at random from all arrays. To model a laminated patch, later, different properties and connectivities will be introduced among the arrays, but keeping the same number of units and connections, to provide for a correct comparison of performance. The 3 arrays will then model supragranular, granular and infragranular layers of the isocortex.

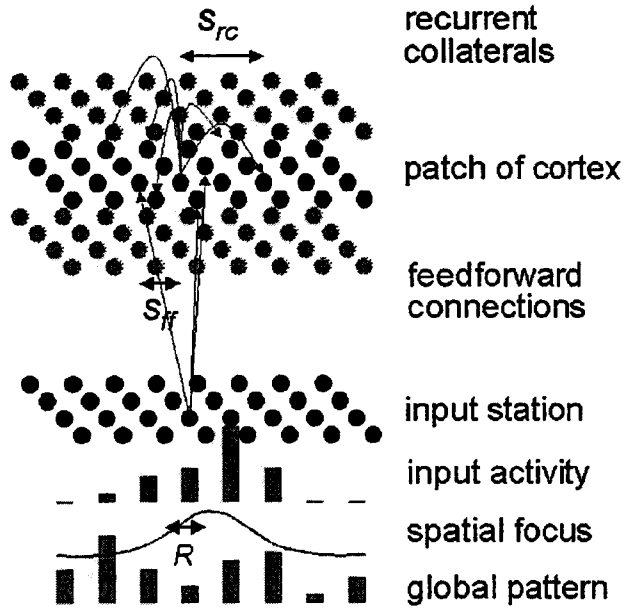

FIG. 1. Scheme of the model patch. Parameters used in the simulations reported: $N \times N = 20 \times 20$, $R = 2$, $S_{rc} = 8$, $S_{ff} = 2 - 8$, $M = 12$, $N_{iter} = 10$. Neural representations are constrained to have sparsity $a = 0.3$ in each layer.

A local pattern of activation is applied to the thalamic units, fed forward to the cortical patch and circulated for N_{iter} time steps along the recurrent connections, and then the activity of some of the units in the patch is read out. To separate out "what" and "where" information, the input activation is generated as the product of one of a set of M predetermined global patterns, covering the entire $N \times N$ input array, by a local focus of activation, defined as a Gaussian tuning function of width R, centered at any one of the N^2 units. The network operates in successive *training* and *testing* phases. In a training phase, each of the possible $M \times N \times N$ activations is applied, in random sequence, to the input array; activity is circulated in the output arrays, and the resulting activation values are used to modify connections weights according to a model associative rule. In a testing phase, input activations are the product of a focus, as for training, by a *partial cue*, obtained by setting a fraction of the thalamic units at their activation in a pattern, and the rest at a random value, drawn from the same general distribution used to generate the patterns. The activity of a population of output units is then fed into a decoding algorithm - external to the cortical network - that attempts to predict the actual focus (its center, p) and, independently, the pattern i used to derive the partial cue. I_i is extracted from the frequency table $P(i, i_d)$ reporting how many times the cue belonged to pattern $i = 1, \ldots, M$ but was decoded as pattern i_d:

$$I_i = \sum_{i,i_d} P(i,i_d) \log_2 \frac{P(i,i_d)}{P(i)P(i_d)} \qquad (1)$$

and a similar formula is used for I_p.

The exact "learning rule" used to modify connection weights was found not to affect results substantially. Those reported here were obtained with the rule

$$\Delta w_{ij} \propto r_j^{post} \cdot (r_i^{pre} - < r^{pre} >) \qquad (2)$$

applied, at each presentation of each training phase, to weight w_{ij}. Weights are originally set at a constant value (normalized so that the total strength of afferents equals that of recurrent collaterals), to which is added a random component of similar but asymmetrical mean square amplitude, to generate an approximately exponential distribution of initial weights onto each unit. r denotes the firing rates of the pre- and postsynaptic units, and $< \ldots >$ an average over the corresponding array. In all the simulations shown, only recurrent weights were modified during training, although making feedforward weights modify as well did not affect substantially the results[2].

Among the several parameters that determine the performance of the network, I set $R << S_{rc}$, and concentrate on S_{ff}, as it varies from $S_{ff} \simeq R$ up to $S_{ff} \simeq S_{rc}$. It is intuitive that if the feedforward connections are focused, $S_{ff} \simeq R$, "where" information can be substantially preserved, but the cortical patch is activated over a limited, almost point-like extent, and it may fail to use efficiently its recurrent collaterals to retrieve "what" information. If the other hand $S_{ff} \simeq S_{rc}$, the recurrent collaterals can better use their attractor dynamics, leading to higher I_i values, but the spread of activity from thalamus to cortex means degrading I_p.

[1]Periodic boundary conditions are used, to limit finit size effects, so the patch is in fact a torus.

[2]To check that the sequential presentation of each local pattern during training was not a crucial factor, I have also presented random combinations of 4 local patterns simultaneously, thereby reducing training time by a factor of 4. Results were unchanged.

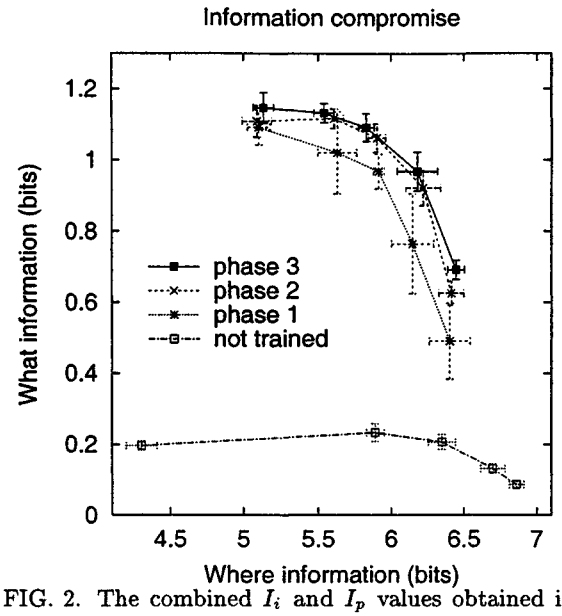

FIG. 2. The combined I_i and I_p values obtained in simulations of the uniform model. The same general curve is obtained with other values for M, and using partial cues of different size. In this and the next Figures, stimuli determined the firing of 40% of the thalamic units, while the remaining 60% had random activity. Nearly asymptotic values are reached after just 3 training phases. Errobars are calculated as the s.d. of the mean of 10 independent runs.

This conflict between I_p and I_i is depicted in Fig.2, which reports their joint values extracted from simulations, as a function of the spread of the afferents, and of the training phase. What is decoded is the activity of all units in the upper array of the patch. Since the patch is not differentiated, however, the other two arrays provide statistically identical information. Further, since information of both the what and where kinds is extracted from a number of units already well in the saturation regime [27], even decoding all units in all 3 arrays at the same time, or only, say, half of the units in any single array, does not alter the numbers of Fig.2 significantly. Before any training occurs, little "what" information can be retrieved; after training (which with these parameters is already asymptotic with 3 epochs) I_i is monotonically increasing with S_{ff}. I_p, instead, decreases with S_{ff}, and as a result one can vary S_{ff} to select a compromise between what and where information, but *not optimise both* simultaneously. This conflict between what and where persists whatever the choice of all the other parameters of the network, although of course the exact position of the $I_p - I_i$ limiting boundary varies accordingly.

Is it possible to go beyond such boundary?

4. Differentiation of a granular layer.

I have explored several modifications of the "null hypothesis" uniform model, to try and find some that could result in a combination of I_p and I_i beyond the limit exemplified in Fig.2. A search of this kind cannot be exhaustive, of course, so I have tried in particular modifications that represent rough models of a granulated patch of cortex. Thus, all changes to the uniform model, at this stage, were designed to model solely the emergence of the granular layer, and not any further aspect of a fully laminated cortex. The values of I_p and I_i obtained with several different simulations, all sharing the same three modifications, but differing in the values of some parameters, are reported in Fig.3. For all values of the parameters, the combined information values alleviate the conflict affecting the uniform model. Of course, other parameter values can be found, that worsen the conflict. It must be stressed, though, that none of these three modifications alone, or in combination with just one of the other two, suffices to cross the boundary. They are required all three together, at least in my experience. The three modifications are:

1. The thalamic afferents to the granular layer are focused, while those to the two pyramidal layers (still the same number, per unit) are diffuse. In the simulations shown, $S_{ff}(IV) = R = 2$, while $S_{ff}(III) = S_{ff}(V) = 8$.

2. The recurrent collateral system of the granular units is severely restricted. In particular, in the simulations reported here, the collaterals originating from layer IV units (and arriving at any layer) are focused ($S_{rc}(IV) = 2 - 3$, while $S_{rc}(III) = S_{rc}(V) = 8$) and non-modifiable by training. Those arriving at layer IV units are fewer in number ($C_{rc}(IV) = 60$ while $C_{rc}(III) = C_{rc}(V) = 150$), thus in fact decreasing the total number of synapses in the laminated model with respect to the uniform one.

3. Model layer IV units follow a non-adaptive dynamics. This is effected in the simulations by making their effect on postsynaptic units, whatever their layer, scale up linearly with iteration cycle. Thus, compared to the model pyramidal units, whose firing rate would adapt over the first few interspike intervals, in reality (but is kept in constant ratio to the input activation, in the simulations), the firing rate of granule units, to model lack of adaptation, is taken to actually increase in time for a given input activation[3].

[3]This could also be taken to model non-depressing short-term plasticity at synapses originating from granule cells, an observation I owe to Prof. Haim Sompolinsky.

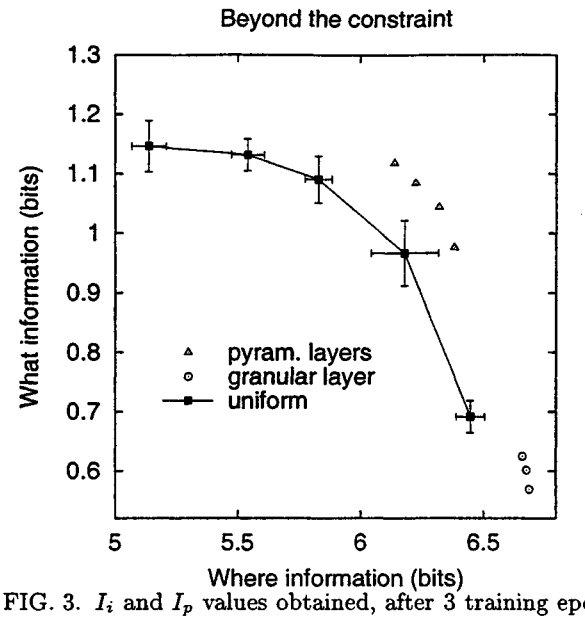

FIG. 3. I_i and I_p values obtained, after 3 training epochs, with the uniform model, and with 4 different parameter choices for the granulated model. Note that the latter is not yet asymptotic after 3 epochs. One of the values decoded from the granular layer (circles) falls outside the graph, at $I_p = 6.66, I_i = 0.51$.

The information values reported in Fig.3 with triangles are decoded from the supragranular layer (layer III). Decoding the activity of layer V, still at this stage statistically identical to layer III, gives the same results. Different values (the circles, with the same parameters as for the layer III triangles) are instead obtained by decoding the activity of layer IV, of course much more biased towards high I_p and low I_i, but still beyond the original constraint. Decoding the activity of all layers simultaneously yields somewhat intermediate values (not shown).

The three modifications, combined, thus produce a slight quantitative advantage in the joint I_p and I_i values that can be read off pyramidal cell activity. The advantage is small, but the model cortical patch used is also tiny ($N \times N = 20 \times 20$), and the expectation is that the difference between uniform and laminated patches would scale up, as the size of the patch reaches realistic values. The crucial factor, in this scaling, is actually the number of recurrent connections each unit receives, which limits the number of global activity patterns which can be stored and retrieved [26]. In the simulations, C_{rc} is some two orders of magnitude below realistic cortical values, but it cannot be made much bigger in a patch of limited size, $N \times N$ - and, on the other hand, scaling up both N^2 and C, which is in principle possible, rapidly makes simulations exceedingly long.

Can we understand the advantage brought about by lamination? The modifications required in the connectivity of layer IV are intuitive: they make granule units more focused in their activation, in register with the thalamic focus, while allowing the pyramidal units, that receive diffuse feedforward connections, to make full use of the recurrent collaterals. What is less intuitive is the requirement for non-adapting dynamics in the granule layer. It turns out that without this modification in the dynamics, the laminated network essentially averages linearly between the performances of uniform networks with focused and with diffuse connectivity, without improving at all on a case with, say, intermediate spread parameters for the connections. This is because the focusing of the activation and the retrieval of the correct identity interfere with each other, if carried out simultaneously, even if the main responsability for each task is assigned to a different layer. Modifying the dynamics of the model granules, instead, enables the recurrent collaterals of the pyramidal layers to first better identify the attractor, i.e. the stored global pattern, to which the partial cue "belongs", and to start the dynamical convergence towards the bottom of the corresponding basin of attraction [28]. Only *later on*, once this process is - in most cases - safely underway, the granules make their focusing effect felt by the pyramidal units. The focusing action, by being effectively delayed after the critical choice of the attractor, interferes with it less - hence, the non-linear advantage of the laminated model.

5. Differentiating infra- from supra-granular connections.

Why does isocortex have pyramidal layers both above and below the granule layer? In the laminated model considered above, the supragranular and infragranular layers are still identical in all their properties, and exactly the same mix of I_p and I_i can be read off the activity of both. In the real cortex, however, the supragranular and infragranular layers differ in several ways. One difference which likely goes back hundreds of millions of years is in their efferent projections. The supragranular layers (denoted here as "layer III", without any commitment about the why, how and when of the further differentiation between layers II and III) project mainly onward, to the next stage of processing[4]. In advanced mammalian species, this means they project to the next cortical areas in the sensory or motor stream [30]. In simpler mammals, and probably in the primordial species, which likely had only one sensory cortical area per modality [31], they project strongly to the medial cortex, associated with multimodal integration and memory [32]. The infragranular layers (denoted here as "layer V", again neglecting the differentiation of V from VI) project mainly backward [33], or subcortically. Among their chief target structures are the very thalamic nuclei from which projections arise to layer IV. It is clear that having different preferential targets would in principle favour different mixes of what and where information. In particular, cortical units that project back to the thalamus would not need to repeat to the thalamus "where" a stimulus is, since this information is already coded, and more accurately, in the activity of thalamic units. They would rather report in its full glory the genuine contribution of cortical processing, that is, the retrieval of identity information. Units that project to further stages of cortical processing, on the

[4]Layer III is also the major source of callosal projections, those to the other hemisphere [29].

other hand, should balance the "what" added value with the preservation of positional information - the mix that we have so far considered optimal for pyramidal units in general.

In addition to the difference in extrinsic projections, the *intrinsic* connectivity of supra- and infragranular layers also differs, although the exact pattern has been explored quantitatively only in some special cases [34]. A useful summary of many ill-determined details is provided by the "canonical" cortical circuit model of Douglas and Martin [35], which describes activity as propagating first to the supragranular and then to the infragranular layers, while being regulated by inhibitory feedback. In fact, while thalamic projections reach the dendrites of units in all three main layers, the subsequent preferential synaptic flow is IV→III→V. I have made an even cruder model of such flow by removing, in a second version of the laminated model, all direct projections from layer IV to layer V, and replacing them with an equal number of projections from layer III to layer V. All other parameters remain as in the first laminated model. With this further modification, layer III becomes the main source of recurrent collaterals [34,36], which are spread out and synapse onto both supra- and infra-granular units and also, to a lesser degree, layer IV units.

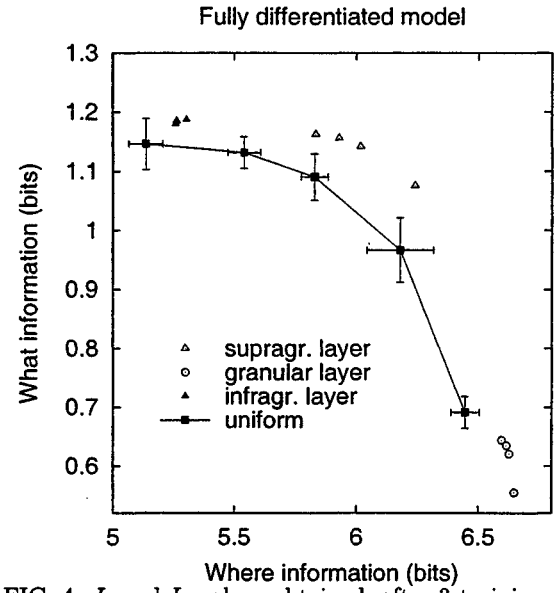

FIG. 4. I_i and I_p values obtained, after 3 training epochs, with the uniform model, and with the 4 parameter choices of Fig.3, but for the fully differentiated model. 3 of the data points for the infragranular layer (black triangles) are nearly superimposed.

The effect of this differentiation can be appreciated by decoding the activity in the three layers, separately, as shown in Fig.4. From layer IV one can extract, as before, a large I_p but limited I_i; from layer III one obtains again a balanced mix, albeit now somewhat biased towards favoring I_i - but a choice of somewhat different parameters would take the balance back to any desired value. From layer V, on the other hand, one can extract predominantly "what" information, I_i, at the price of a rather reduced I_p content. Thus, the further connectivity change (which leaves total synaptic numbers unaltered), by effectively reducing the coupling between granular and infragranular layers, has made the latter optimize "what" information, while neglecting "where" information, of limited interest to their target structures.

6. Discussion and falsifiability.

I propose that a small quantitative advantage in relaying combined positional and identity information may have driven the differentiation of neocortical layers in mammals.

The proposal, while in line with speculations arising of traditional comparative neuroanatomy, is supported by the simulation of an offendingly crude neural network model - a methodology which requires justification. Connectionist models have a reputation for being adaptive, which often means they can be designed *ad hoc* to demonstrate the validity of whatever hypothesis. Indeed, the plethora of parameters which have to be specified in even a simplified neural network model, like the one simulated here, is so large as to make exhaustive analysis impossible, and independent validation difficult. In the case of my simulations, not only the analysis of parameter space, but even most of the details of the model could not be reported here, for lack of space, and will be described in full elsewhere[5]. The truly important elements, however, are the mutual constraint between relaying where information and retrieving what information, evident in the uniform model, and the quantitative comparison with a laminated model with no more units and/or connections. The what/where conflict is manifest also in rather different models, like those developed independently by Hamish Meffin [37], and, essentially, it requires only the separate measurability of I_i and I_p to be demonstrated, whatever the remaining details of the model. The quantitative comparison is taken to be fair, here, since the laminated model remains identical to the uniform original, except for the modifications discussed, and in particular has the same number of units. The overall number of synapses in fact decreases slightly, to allow layer IV units, which receive only 2/5 of the original recurrent collaterals, to be more influenced by thalamic inputs in the laminated version. In the real cortex, which devotes most of its volume to synapses [17], it is likely that synaptic density per mm^2 is a true constraint to evolutionary expansion; it is also possible that the number of synaptic inputs on a pyramidal cell is limited by biophysical constraints, to keep the effective electrotonic length of the dendrites short, and allow efficient integration by the cell. Whether the number of cells per cortical mm^2 may also be limited, is less clear. Keeping it fixed, in any case, ensures a conservative comparison, which makes the advantage of lamination controlled for trivial effects.

The small advantage of the laminated patch is expected to scale up, as mentioned above, when the model, and in particular synaptic numbers, are scaled up to real-

[5]The code used may be obtained from my website `http://www.sissa.it/~ale/limbo.html`

istic values. It should be considered, however, that even a slight quantitative advantage may be selected for, once replicated over millions of sensory experiences per individuals, and over millions of generations in the course of mammalian evolution. Such a quantitative advantage can obviously be demonstrated only with computer simulations, which are precise and can also be replicated millions of times. It remains inaccessible to experimental observation, either *in vivo* or *in vitro*, even if it where possible to devise preparations that approximated laminated and uniform cortical patches with similar quantitative characteristics.

What are then the predictions arising from the proposal, that could be checked experimentally? Essentially, differences in the information content of the activity of populations of cells in different layers. With an appropriate experimental design, I_p and I_i can be measured *in vivo* from populations of tens of units [27] recorded in well identified layers. While the relative values of I_p and I_i depend on the design and are not comparable, the model does predict that, very much as in Fig.4, when separate measures are extracted from populations of equal size, $I_p(V) \ll I_p(III) \leq I_p(IV)$, while $I_i(V) \geq I_i(III) \gg I_i(IV)$. The differences in I_i should be manifest in cortical areas crucial in the processing of the stimuli to be discriminated, and the testing should involve the use of rather noisy stimuli (partial cues to retrieval). In addition, the time course of $I_p(III)$ is expected to be delayed with respect to that of $I_i(III)$.

How does my proposal relate to alternative accounts of the significance of neocortical lamination? Essentially, it does not interfere nor cooperate with the few accounts that, to my knowledge, have been proposed. For example, the 'RULER' model [38] emphasizes the *dynamics* of activation in the different layers, in line with the canonical model of Douglas *et al.* [35], but it does not attempt to really quantify function, or performance. The present model, which is extremely simplified in its dynamics, should be entirely compatible with a more accurate dynamical description. A number of papers have been produced by Stephen Grossberg and collaborators [39] which as a whole relate neural interactions between the various layers to mechanisms of visual perceptions, e.g. to promote the grouping together of V1 cells with similar orientation and disparity selectivity, or of V2 cells that represent similar edges, texture or shading. The mechanisms described are fairly complex and difficult to assess with quantitative comparisons between laminated and uniform models. While it remains possible that some of these mechanisms might be specific implementations of my generic account, or at least might be compatible with it, the perspective is clearly very different. It seems to me difficult to disentangle, from the sophisticated mechanisms that have evolved in visually advanced species, such as cats or monkeys, the primitive ones that may have been associated with the emergence of lamination, hundreds of millions of years before. A simpler strategy seems to be the one pursued here, of considering generic aspects of sensory information processing, pertinent to each topographic modality, and which lend themselves easily to accurate quantification, at least in terms of computer models.

Acknowledgments.

I am indebted to Hamish Meffin, for enbarking with me on this project, although our routes parted half-way, and to Mathew Diamond and Israel Nelken, for considering experimental verifications. Partial support from the Human Frontier Science Programme grant RG0110/1998-B.

---

[1] Carroll, R.L. (1988) Vertebrate Paleontology and Evolution (W H Freeman & Co., New York).
[2] Jerison, H.J. (1990) In Cerebral Cortex, vol. 8A: Comparative Structure and Evolution of Cerebral Cortex, eds.Jones, E.G. & Peters, A. (Plenum Press, New York), pp.285-309.
[3] Wilson, E.O. (1975) Sociobiology. The New Synthesis (Harvard Univ. Press, Cambridge, MA).
[4] Diamond, I.T. & Hall, W.C. (1969) Science 164:251-262.
[5] Finlay, B.L. and Darlington, R.B. (1995) Science 1268: 1578-1584.
[6] Ulinski, P.S. (1990) In Cerebral Cortex, vol. 8A: Comparative Structure and Evolution of Cerebral Cortex, eds.Jones, E.G. & Peters, A. (Plenum Press, New York), pp.139-215.
[7] Amaral, D.G., Ishizuka, N. & Claiborne, B. (1990) Prog. Brain Res 83:1-11.
[8] Diamond, I.T., Conley, M., Itoh, K. & Fitzpatrick, D. (1985) J. Comp. Neurol. 242:610.
[9] Marr, D. (1971) Phil. Trans. R. Soc. Lond. B 262:24-81.
[10] Treves, A. & Rolls, E.T. (1992) Hippocampus 2:189-199.
[11] Lassalle, J.-M., Bataille, T. & Halley, H. (2000) Neurobiol. Lear. Mem. 73:243-257.
[12] Allman, J. (1990) In Cerebral Cortex, vol. 8A: Comparative Structure and Evolution of Cerebral Cortex, eds.Jones, E.G. & Peters, A. (Plenum Press, New York), pp.269-283.
[13] Lende, R.A. (1963) Science 141:730-732.
[14] Donoghue, J.P., Kerman, K.L. & Ebner, F.F. (1979) J. Comp. Neurol. 183:647-666.
[15] Haberly, L.B. (1990) In Cerebral Cortex, vol. 8B: Comparative Structure and Evolution of Cerebral Cortex, eds.Jones, E.G. & Peters, A. (Plenum Press, New York), pp.137-166.
[16] Abeles, M. (1991) Corticonics: Neural Circuits of the Cerebral Cortex (Cambridge Univ. Press, Cambridge).
[17] Braitenberg, V. and Schüz, A. (1991) Anatomy of the Cortex (Springer-Verlag, Berlin).
[18] Kaas, J.H. (1982) In Contributions to Sensory Physiology, vol. 7 (Academic Press, New York) pp.201-240.
[19] Krubitzer, L. (1995) Trends Neurosci. 18:408-417.
[20] Jones, E.G. (1998) In Consciousness: At the Frontiers of Neuroscience. eds Jasper, H.H., et al. (Wiley-Liss, New York).
[21] Erickson, R.P., Hall, W.C., Jane, J.A., Snyder, M. & Diamond, I.T. (1967) J. Comp. Neurol. 131:103-130.
[22] Rauschecker, J.P., Tian, B. & Hauser, M. (1995) Science 268:111-114.
[23] Hopfield, J.J. (1982) Proc. Natl. Acad. Sci. USA 79:2554-2558.
[24] Amit, D.J. (1995) Behav. Brain Sci. 18:617-657.
[25] Whitfield, I.C. (1979) Brain Behav. Evol. 16:129-154.

[26] Rolls, E.T. & Treves, A. (1998) Neural Networks and Brain Function (Oxford Univ. Press, Oxford).
[27] Treves, A. (2001) In Handbook of Biological Physics, vol. 4: Neuro-Informatics and Neural Modelling, eds Moss, F. & Gielen, S. (Elsevier, Amsterdam) pp. 825-852.
[28] Amit, D.J. (1989) Modelling Brain Function (Cambridge Univ. Press, New York).
[29] Innocenti, G.M. (1986) In Cerebral Cortex, vol. 5: Sensory-Motor Areas and Aspects of Cortical Connectivity, eds.Jones, E.G. & Peters, A. (Plenum Press, New York), pp.291-353.
[30] Barbas, H. & Rempel-Clower, N. (1997) Cereb. Cortex 7:635-646.
[31] Rowe, M. (1990) In Cerebral Cortex, vol. 8B: Comparative Structure and Evolution of Cerebral Cortex, eds.Jones, E.G. & Peters, A. (Plenum Press, New York), pp.263-334.
[32] Gloor, P. (1997) The Temporal Lobe and Limbic System (Oxford University Press, New York).
[33] Batardiere, A., Barone, P., Dehay, C. & Kennedy, H. (1998) J. Comp. Neurol. 396:493-510.
[34] Nicoll, A. & Blakemore, C. (1993) Neural Comput. 5: 665-680.
[35] Douglas, R.J., Martin, K.A.C. & Whitteridge, D. (1989) Neural Comput. 1:480-488.
[36] Yoshioka, T., Levitt, J.B. & Lund, J.S. (1992) J. Neurosci. 12:2785-2802.
[37] Meffin, H. (2001) Advances in Neural Information Processing, in press.
[38] McComas, A.J. & Cupido, C.M. (1999) Clin. Neurophysiol. 110:1987-1994.
[39] Grossberg, S. (1999) Spat. Vis. 12: 163-185.