3 Spatial Cognition, Memory Capacity, and the Evolution of Mammalian Hippocampal Networks

Alessandro Treves

Beyond the Grid, There Is Not Much Space

The discovery of grid cells in the medial entorhinal cortex (MEC) of the rat (Fyhn et al. 2004) and of the precise triangular pattern of their firing fields (Hafting et al. 2005) requires a substantial reformulation of the questions relating to spatial cognition. It now appears, more clearly than before, that spatial computations per se are largely performed by the rat brain before the hippocampus is ever accessed, and culminate in a sort of universal map of allocentric space, in MEC layer II. Only a portion of the EC participates in such a map, which is applied and used irrespective of context, but in combination with context-specific signals that determine the activity of other parts of entorhinal cortex. The hippocampus operates on the universal map and on context-specific signals to create context-specific metric representations of space, which are stored in memory. The capacity of the hippocampus to rapidly switch between the representations of different contexts is illustrated by hippocampal global "remapping," i.e., the transition to new, unrelated arrangements of place fields by the same population of recorded cells, after suitable behavioral manipulations, and without concurrent MEC remapping (Fyhn et al. 2007). Consequently, understanding the circuitry of the hippocampus crucially involves understanding this capacity for decorrelating spatial representations, at least in rodents. It could well be that in other species complex memories of a less spatial nature take a more prominent role, in which case it would be even more appropriate to approach hippocampal decorrelation and memory processes at an abstract level, and independent of the possibly species-specific spatial processes so finely investigated with the rat model (see also chapter 2 of this volume, by Lucia F. Jacobs).

Which approaches can take us beyond a mere functional description of the role of different networks in the brain, and lead us to understand, in evolutionary terms, their design principles? In recent years I have studied three apparently disparate topics from a computational viewpoint: the lamination of the sensory cortex, the differentiation into subfields of the mammalian hippocampus, and the neuronal dynamics that might underlie the faculty for language in the human frontal lobes. These studies share a common perspective: they all discuss the evolution of cortical networks in terms of their computations, quantified by simulating simplified formal models. They all dwell on the interrelationship between qualitative and quantitative change. Finally, they all include, as a necessary ingredient of the relevant computational mechanisms, a simple feature of pyramidal cell biophysics: firing-rate adaptation. In this chapter I formulate this general viewpoint, which does not usually find space in individual papers, and then I focus on the computational approach to hippocampal network design, seen in the context of the other two problems.

Looking at the Past Through a Spin Glass

To approach each of the three problems, I have used the simulation of drastically simplified network models as the primary tool for analysis. Although the details of the models used were specific and were adapted to the problem being considered, the underlying approach has been similar across studies, and this is what I want to briefly discuss first.

An assumption motivating my approach is that the most important steps in the evolution of the nervous system are those that address *computational* demands, demands that are part of the "job specification" of the brain as an information-processing system, rather than steps that address, say, physiological or anatomical constraints. Among genuine information-processing problems, one that has been quantified through the use of formal models is the limit on the storage of memories that is imposed by the connectivity of a system of neuron-like units. Considering this limit is partly motivated by the observation that most gray-matter volume appears to be devoted to synaptic contacts (Braitenberg and Schüz 1991), as if the cortex had evolved to maximize connectivity and ultimately memory storage. The mathematical procedures that have been used to obtain a proper quantification of the relation between connectivity and memory were originally developed to analyze the physics of a class of materials known as spin glasses (see, e.g., Amit 1989). Spin glasses are endowed with interactions that can be characterized as disordered and hence as interfering with each other, somewhat as, in a neural network, distinct memory representations interfere with each other at retrieval. Although spin glasses have nothing deeper in common with memory systems than this analogy and the mathematical procedures useful in analyzing them, the effectiveness and generality of these procedures have led some of us to approach many information- processing problems by relying on the analysis of spin glasses as a basic paradigm. Unwrapped from its technicalities, the spin-glass approach reduces essentially to the idea that cortical systems face a crucial connectivity constraint on extensive memory storage, that the constraint results from interference among memories, and that to analyze such interference we can borrow techniques from statistical physics.

The three problems I have considered are all, to some extent, spin-glass problems in disguise.

The Phase Transition That Made Us Mammals

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 300 million years ago (Carroll 1988). Perhaps mammals emerged as a fully differentiated but still rather homogeneous class out of the third-to-last of the great extinctions, in the Triassic period, with their explosive diversification occurring much later (Bininda-Emonds et al. 2007). 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; Jerison 1990) and of presumed behavioral patterns (Wilson 1975), the main line of evidence is the comparative anatomy of present-day species (Diamond and Hall 1969). Among a variety of quantitative changes in the relative development of different structures—changes that, by tuning the expression of specific genes (Mallamaci and Stoykova 2006) have been extended, accelerated, and diversified during the entire course of mammalian evolution (Finlay and Darlington 1995; Barton 2007)—two major qualitative changes in the forebrain stand out: 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 change, it is the medial pallium (the medial part of the upper surface of each cerebral hemisphere, 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, which loses both its continuity with the rest of the cortex at the hippocampal sulcus and its projections to the dorsolateral cortex (Ulinski 1990). 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, which now project, as a sort of preprocessing stage, to the pyramidal cells of field CA3 (Amaral, Ishizuka, and Claiborne 1990). In the second change it is the dorsal pallium (the central part of the upper surface) that reorganizes internally, in areas that process topographic modalities, 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 classified as 6, in some cases with specialized sublayers; see Yamamori and Rockland 2006). A prominent step in lamination is granulation, whereby the formerly unique principal layer of pyramidal cells is split by the insertion of a new layer of excitatory, but intrinsic, granule cells, in between the pyramidal cells of the infragranular and supragranular layers. This is layer IV, where the main ascending inputs to cortex terminate (Diamond et al. 1985).

Lamination May Reconcile Memory with Topography

I have formulated a hypothesis (Treves 2003) that accounts for granulation, and for the differentiation between supra- and infragranular pyramidal layers, as advantageous to support fine topography in the sensory maps that mammals have evolved, over and beyond the gross topography that limits the usefulness of sensory maps in reptiles. 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. Memory patterns can be stored on recurrent collaterals in the cortex, and such memory can help substantially in the analysis of current sensory input. The effective use of recurrent collaterals, because of the “spin-glass” limit on memory storage load, requires afferent projections to the cortex that are spread over a large patch; whereas the precise localization of a stimulus on the sensory map requires narrowly focused afferents (see Treves 2003 for the complete argument; Roudi and Treves 2006 for the analytical treatment of a single-layer model). The simulation of a simplified network model demonstrates that a nonlaminated patch of cortex with a single characteristic spread of afferent connections must compromise between transmitting “where” information or retrieving “what” information. The differentiation of a granular layer affords a quantitative advantage by allowing focused afferents to the granular units together with widespread afferents to pyramidal units. For this purely anatomical differentiation to be effective, however, it must be accompanied by a physiological differentiation: pyramidal units must adapt their firing—that is, decrease their response to steady inputs—much more than granular units. With this further difference, the pyramidal layers can select the correct attractor for memory retrieval before the granular layer, which adapts less, partially takes over the dynamics, and focuses activity on the cortical spot that most accurately reflects the position of the sensory input.

Adaptation thus effectively separates out in time, albeit only partially, two information-processing operations that occur in different spaces: the retrieval of memories in the abstract space of attractors and the accurate relay of stimulus position in the physical space of the cortical surface. The advantage of the differentiation is quantitatively minor (see figure 3.1). My hypothesis is that a major qualitative step, the transition from a simpler paleocortex to a more elaborate isocortex, came about just in order to gain a few percent more bits in the average combined value of “what” and “where” information.

The Phase Transition That Made Us Human

Our lineage is estimated to have radiated away genetically from those of other great apes perhaps 5 million to 6 million years ago. Functionally, a number of lines of evidence point at a stage of accelerated change and complexification in human behavior, a so-called cognitive revolution, taking place much later, perhaps 50,000 years ago. In terms of the organization of the nervous system, no salient qualitative trait distinguishes us from our

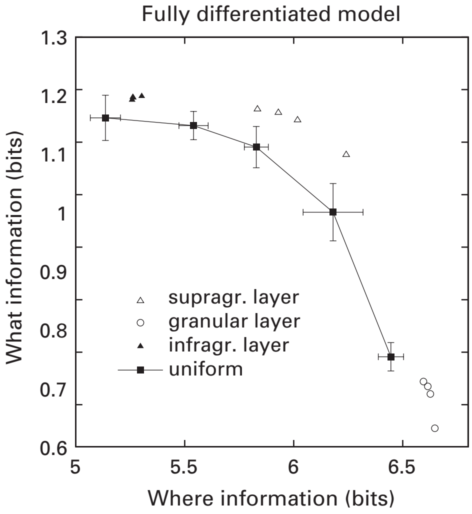

Figure 3.1
Combination of what and where information that can be extracted from neural activity in model cortical patches of different architectures. The "uniform" model is made up of three statistically identical layers of units, each of which produces the same what-where mix at some location on the solid line, depending on the spread of afferent inputs. The "differentiated" model includes three layers with distinct connectivity and adaptation properties, each of which affords a different what-where mix, but always beyond the limit reached by the uniform model. Simulation details in Treves (2003).

closest cognates, such as chimps or gorillas. The only clear structural pattern is one of quantitative change: with respect to other primates—but not necessarily with respect to all other mammals (Goffinet 2006)—an increase is observed in some key parameters of cortical extension, arealization (Krubitzer and Huffman 2000), and connectivity (Elston 2000). It seems unlikely, therefore, that a new gene may have triggered the development of uniquely human capacities without apparently inducing any detectable change of design in the brain. Yet the scientific community has been reluctant to abandon the expectation, naively raised by popular media, of a quick and ready-to-use solution to the question of what makes us human. Many respectable scientists have continued to harbor the hope that—if not a gene, if not a magic molecule—at least a dedicated piece of neural circuitry may be found that could explain, for example, the human faculty of language.

Understanding the neural basis of higher cognitive functions such as those involved in language requires in fact a shift from a localization approach to an analysis of network operation. Localization approaches have run their course, and they have highlighted a substantial continuity between the cortical areas most directly implicated in language functions in the human brain and homologue areas in other primates. Finding out where language "is" has not provided a shred of a clue as to how it came about. Hauser and colleagues, in a recent proposal (2002), point instead to *infinite recursion* as the core process involved in several higher functions, including language, forcefully arguing for the hypothesis that the roots of language may be in a new process rather than in a new structure. The proposal offers a way out of the reductionist cul-de-sac and it challenges cortical-network theorists to describe network behavior that could subserve infinite recursion.

Building on a variant of the notion that language may have evolved out of the semantic and procedural memory systems (Ullman 2001) I have been exploring the hypothesis that a capacity for infinite recursion may be associated with the natural adaptive dynamics of large semantic associative networks (Treves 2005). I have used a network of Potts (multistate) units to simulate a semantic memory system distributed over many cortical modules, and I have tested its joint ability to both retrieve a semantic memory based on a partial cue and, subsequently, when deprived of further inputs, also to follow a latching dynamics in attractor space, jumping from one memory to the next with structured transition probabilities. While the retrieval ability is limited by an appropriate variant of the spin-glass constraint (first considered by Kanter 1988), the latching ability requires a sufficient density of attractors. Since the spin-glass constraint limits the number of attractors proportionally to the connectivity (Kropff and Treves 2005), the joint ability for retrieval and latching can be realized only once the connectivity of the modular system becomes, in evolution, sufficiently extensive. At that point, after a kind of percolation-phase transition, the system is both able to retrieve and to support structured transition probabilities between global network states. The crucial development endowing a semantic system with a nonrandom dynamics would thus be an increase in connectivity, perhaps to be identified with the dramatic increase in spine numbers recently observed in the basal dendrites of pyramidal cells in human and Old World monkey frontal cortex (Elston 2000). Once again the crucial step in the argument is a quantitative analysis based on network simulations, which spin-glass mathematical methods promise to consolidate, to describe a phase transition that could not be accessed with mere qualitative reasoning.

The Differentiation of the Hippocampus

Focusing now on the hippocampus, one may ask, what is the evolutionary advantage, for mammals, brought about by the changes mentioned above, in its internal organization? Since the seminal paper by David Marr (1971), and well before awareness developed among modelers of the evolutionary specificity of hippocampal organization (Treves et al. 2008), attempts to account for its remarkable differentiation into three main subfields have been mostly based on the computational analysis of the role of the hippocampus in memory. With the simultaneous discovery of place cells (O'Keefe and Dostrovsky 1971), the rodent model seemed to point at a special hippocampal role for spatial representation and spatial memory. Although an accumulating body of evidence has suggested that hippocampal activity may not be exclusively related to space (Eichenbaum 2000), the prevalence of spatial correlates in the rat has encouraged speculations on the evolution of the hippocampus based on spatial function. The hippocampus, however, is important for spatial memory also in birds (Bingman and Jones 1994; Clayton and Krebs 1995; Clayton et al. 2003; Bingman and Sharp 2006). The avian and mammalian hippocampi are structurally very different, with birds perhaps having stayed close to their reptilian progenitors in this respect, and mammals having detached the dentate gyrus from Ammon's Horn, as mentioned above and further discussed by Treves and colleagues (2008). A reasonable hypothesis may then be that the new mammalian design somehow enhances the capability of the hippocampus to serve as a memory store, perhaps with the nuance of a prevailingly spatial memory store.

It is plausible that the primitive cortical tissue in early reptile-like ancestors of both mammals and birds was rich in recurrent collaterals, much like region CA3 in the modern mammalian hippocampus. Simplified models show how a recurrent network can naturally retrieve distributions of activity from partial cues as an autoassociative memory (Hopfield 1982), provided the synapses on the recurrent connections among its pyramidal cells are endowed, as likely was the case for primitive cortex, with associative, "Hebbian," plasticity, such as that based on NMDA receptors (Collingridge and Bliss 1995). That cortex can then be conceptualized as having operated, at least, as a content-addressable memory for distributed activity patterns—provided it had an effective way of distinguishing its operating modes. A generic problem with associative memories based on recurrent connections is to distinguish a storage mode from a retrieval mode. To be effective, recurrent connections should dominate the dynamics of the system when it is operating in retrieval mode. While storing new information, instead, the dynamics should be primarily determined by afferent inputs, with limited interference from the memories already stored in the recurrent connections, which should, however, modify their weights to store the new information (Treves and Rolls 1992).

Distinguishing Storage from Retrieval

The most phylogenetically primitive solution to effect the dual operating mode is to use a modulator that acts differentially on the afferent inputs (originally, those arriving at the apical dendrites) and on the recurrent connections (predominantly lower on the dendritic tree). Acetylcholine (ACh) can achieve this effect, exploiting the orderly arrangement of pyramidal cell dendrites in the cortex (Hasselmo and Schnell 1994). Acetylcholine is one of several very ancient neuromodulating systems, well conserved across vertebrates, and it is likely that it operated in this way already in the early reptilian cortex, throughout its subdivisions. Mike Hasselmo has emphasized this role of ACh in memory with a combination of slice work and neural network modeling (Hasselmo et al. 1995, 1996). This work has been focused on the hippocampus, originally the medial wall, and on piriform cortex, originally the lateral wall. The proposed mechanism, however, has no reason to be circumscribed to these regions, and it could well operate across cortical systems involved in memory storage.

One flaw of an ACh-based mechanism is that it requires an active process that distinguishes storage from retrieval periods, and regulates Ach release accordingly. In the hippocampus, however, it appears that mammals have devised a more refined expedient to separate storage from retrieval, which can efficiently perform both functions also in a passive mode: inserting a preprocessor before the CA3 memory network. The preprocessor should instruct which units in CA3 should fire in a new distribution of activity to be stored as the memory trace of a new item to be remembered. As a simplest model, one can think of a preprocessing network without recurrent connections, which simply forms new arbitrarily determined representations on the fly, and through a system of one-to-one connections ("detonator" synapses) imposes these new representations onto CA3 (McNaughton and Morris 1987). In fact, the one-to-one correspondence is not needed: what enriches the representation to be stored of meaningful content, against the interference of recurrent connections, is just a system of sparse and strong connections from a sparsely coded feedforward network. Developing the preprocessor notion, we have proposed a quantitative estimate of the amount of new information that could be encoded in CA3 representations with different input systems (Treves and Rolls 1992; see figure 3.2).

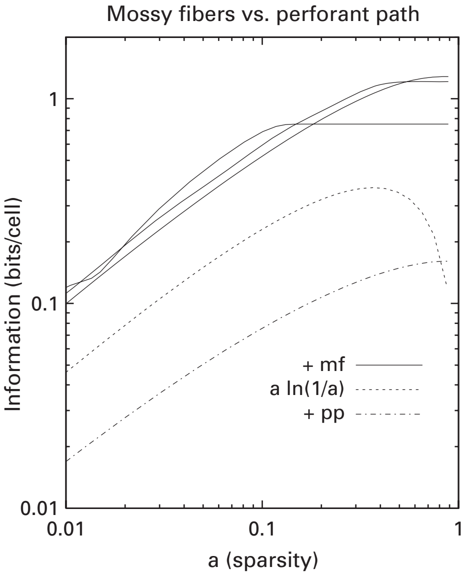

Figure 3.2
Information per unit in a model CA representation, as a function of its sparsity. The dashed line is the amount that can be associatively retrieved; the dash-dotted line is what can be stored by inputs with the characteristics of the perforant path to CA3, and the three solid lines are three estimates of what can be stored through mossy fiber inputs (the estimates differ in the sparsity of the dentate). Analytical derivation in Treves and Rolls (1992).

The argument is based on the "quasi-theorem," which has never been satisfactorily proved but empirically holds true, that an associative memory can hold up to $I/(NC) \approx$ 0.2–0.3 bits of information per synapse, where N is the number of units and C the average number of connections, or synapses, per unit. Since the storage capacity, or maximum number of discrete patterns of activity that can be individually retrieved, is estimated as $p_c \approx 0.2\text{–}0.3\ C/[a \ln(1/a)]$, where a is the sparsity of the stored representation (see Treves and Rolls 1991), information-theoretical efficiency requires that each such representation should contain at least roughly $i \approx N\ a \ln(1/a)$ bits of new information. Per unit, this is the amount of information of a noiseless binary variable (in the sparse $a \ll 1$ regime and apart from the ln(2) factor). Thus, efficient storage requires that CA3 pyramidal units be as informative about new contents as, roughly, binary units can be. The challenge for afferent inputs is to prevail over the recurrent connections, which do not impart new contents to a pattern of activity to be stored. Figure 3.2 shows that this challenge can be met by afferent inputs with the characteristics of the mossy fibers, but not by those with the characteristics of the perforant path to CA3 (Treves and Rolls 1992).

Mapping Continuous Attractors onto Discrete Memories

The argument above has been worked out for the case of discrete memory items, which can be taken as a model of episodic memory. Initially, in fact, the neural network approach, aiming at quantifying the capacity of associative memories, has been formulated in terms of fully connected recurrent architectures and discrete memory states, conceived—in the limit of no fluctuations—as points in the multidimensional space in which each component corresponds to the firing rate or in general to the activity of one unit (Hopfield 1982). This formulation, which was the starting point for physicists interested in applying powerful mathematical analysis techniques, had been preceded by the more rudimentary analysis of David Marr. Marr also thought in terms of discrete memory states, and had guessed the importance of recurrent collaterals, a prominent feature of the CA3 subfield (Amaral et al. 1990), even though his own model was not really affected by the presence of such collaterals, as shown later (Willshaw and Buckingham 1990). Although the paper by Marr was nearly simultaneous with two of the most exciting experimental discoveries related to the hippocampus, that of place cells (O'Keefe and Dostrowski 1971) and that of long-term synaptic potentiation (Bliss and Lømo 1973), for a long time it did not seem to inspire further theoretical analyses, with the exception of an interesting discussion of the collateral effect in a neural network model (Gardner-Medwin 1976). One factor was probably the mathematical "technology" available to Marr, inadequate to really investigate his models quantitatively. Marr himself become disillusioned with his youthful enthusiasm for unraveling brain circuits, and in his mature years took a much more sedate, and less neural, interest in vision. From the 1987 paper by McNaughton and Morris, however, an increasing number of other investigators rediscovered the young Marr, and tried to elaborate those ideas in order to understand the operation of hippocampal circuits. Edmund Rolls (1989) and others again emphasized the crucial role probably played by the CA3 recurrent collaterals and made explicit the relation to the auto-associative memory networks studied quantitatively by the physicists (Amit et al. 1987). In establishing the relation, the salient spatial character of hippocampal memory correlates was provisionally neglected, to take advantage of the formal models based on discrete attractor states.

As a matter of fact, an autoassociator may subserve both the storage of discrete memories as point-like attractor states or of more complex memories—for example, synfire chains (Abeles 1991), which can be individually distinct and discrete or organized in arbitrary branching patterns—or continuous attractors, when network dynamics converges to fixed points that are continuously arranged on some manifold in the high–dimensional activity space. Simple examples of continuous attractors are present in models of orientation selectivity by horizontal interaction in visual cortex (Sompolinsky and Shapley 1987) or of the head direction system (Skaggs et al. 1995). These models do not store information in long-term memory, and in the continuum limit their fixed points comprise a single (in these particular cases, 1D) manifold. Samsonovich and McNaughton's multiple-chart model (1997) demonstrated instead, in the context of a model for path integration, how one could conceive of fixed points organized in multiple 2D continuous manifolds, each of which maps the animal position in a distinct environment. Exploration of a new environment leads to the formation of a new chart (*ab initio,* or using some prewired connectivity; it may be difficult to distinguish the two possibilities). The question then arises of how many charts a given recurrent network can hold in long-term memory.

The storage capacity of a multichart recurrent autoassociator was analyzed by Battaglia and Treves (1998), who extracted a simple rule of thumb for assessing the memory load of a chart. A chart that maps a finite environment onto the activity of place-cell-like units is equivalent, capacitywise, to as many discrete attractor states as there are locations, in the environment, for which the activity vectors are pairwise decorrelated. If the 2D environment is represented by place-cell-like units, which are quiescent outside their place field, the decorrelation radius is roughly the radius of the typical place field, which is itself proportional to the linear size of the environment times the square root of the sparsity of the neural representation. Thus, if, say, some dozen typical CA3 fields "fit", once properly juxtaposed, in a typical rat recording box, the memory load of the chart corresponding to that box is roughly equivalent to a dozen discrete memories of equal sparsity. The number of such charts, or distinct environments that can be held simultaneously in the network, is limited by the critical value $p_{charts} \approx 0.1\ C\ /\ \ln(1/a)$ (see figures 1 and 2 in Battaglia and Treves 1998). The apparent paradox that fewer charts can be stored if they are sparser (a lower a parameter makes the denominator larger) can be understood by considering that sparser activity, in a large net, leads to better spatial resolution, and hence requires more discrete fixed-point attractors to cover, as effectively smaller tiles, the whole environment. This chart capacity again respects the unproven associative memory theorem mentioned earlier, in that the maximum amount of information that can be retrieved per synapse is about 0.15 bits, as shown in figure 5 of Battaglia and Treves (1998).

The Dentate Gyrus as a Chart Preprocessor

This mapping quantifies the retrieval capacity for charts and opens the way for once more investigating the issue of whether enough new information can be stored in each representation to fully exploit the network capacity for information retrieval. In other words, a quantitative analysis of information storage in a model CA3 network, operating with and without dentate gyrus, would be needed to assess again any information-theoretical advantage in forming new representations. Unfortunately, the very 2D nature of charts makes a simplified mathematical analysis of information storage like the one in Treves and Rolls (1992) not applicable, because neighboring locations on one new chart generate correlated activity that cannot be easily dissected from the interfering correlations with other, unrelated, charts. Mathematical tools based on the newly revealed spatial representation in the dentate are being developed (Treves, Cerasti, and Papp 2008); meanwhile I have reported simplified simulations (Treves 2004). Within their limits, the simulations confirm the essential role of the inputs from the dentate gyrus to CA3 in guiding the learning of a new chart (see figure 3.3).

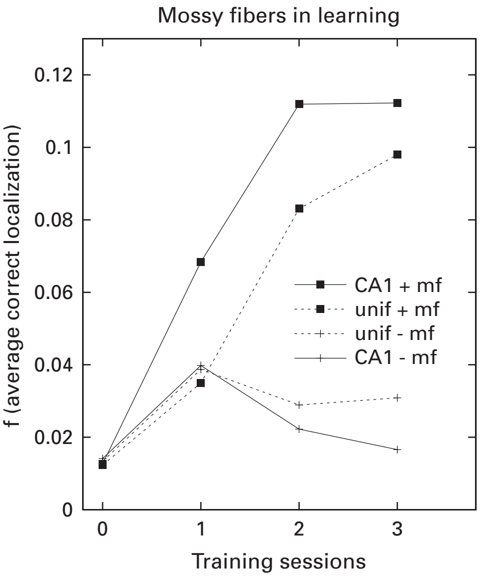

Figure 3.3
Localization accuracy as a function of training session in hippocampal network models with and without the differentiation between CA3 and CA1, and with and without mossy fiber inputs. In the differentiated model, the position of the virtual rat is decoded from the CA1 representation. With reference to figure 3.4, cue size is Q = 0.2. Other details in Treves (2004).

It is worth noting that the beneficial forcing effect of mossy fiber inputs to CA3 is even more salient when assessed indirectly in the information content or localization accuracy afforded by representations in CA1, whose units are only indirectly influenced by dentate gyrus activity.

These simple simulations, while supportive of a specific role of the dentate gyrus in information storage, did not take into consideration two major recent findings. One concerns the differences emerging between CA3 and CA1 in the representation of similar, correlated environments; these will be discussed further. The other is the novel observation that the activity of dentate units, previously deemed to be very sparse, seems to be concentrated on a relatively small fraction of newly generated granule cells (Ramirez-Amaya et al. 2006), which are biophysically indistinguishable from older neurons (Laplagne et al. 2006) but appear to "take care" of representing new information more than older, and perhaps already committed, neurons. New recording experiments (Leutgeb et al. 2007) may help clarify how space is coded by dentate granule cells.

At any rate, the crucial prediction of the argument based on the analysis of discrete memories, if applied to charts, is that the inactivation of the mossy fiber synapses should impair the formation of new charts, but not the retrieval of previously stored ones. This prediction has recently been supported at the behavioral level (Lassalle et al. 2000): mice with a temporary inactivation supposedly selective for the mossy synapses were impaired in finding the hidden platform in a Morris water maze, but not if they had learned its location the previous week. A consistent result was more recently obtained in rats with a different, irreversible procedure of selective lesions, and using indicators that only very approximately dissociate storage from retrieval (Lee and Kesner 2004): the strong double dissociation found between perforant path and dentate lesions is remarkable, given the overlapping nature of the behavioral measures. While waiting for neurophysiological experiments to test the prediction at the neural level, the tentative conclusion from behavioral tests in rodents is that indeed the dentate gyrus may have evolved in order to facilitate the storage of new information in the recurrent CA3 network. If validated, this hypothesis suggests that a quantitative information-theoretical advantage may have favored a qualitative change, such as the insertion of the dentate gyrus in the hippocampal circuitry.

CA1 as a Cleanup Device?

The DG argument does not itself address the CA3-CA1 differentiation, which is equally prominent in the mammalian hippocampus. If DG can be understood as a CA3 preprocessor, perhaps CA1 should be understood as a CA3 postprocessor. In reptiles, CA3 and CA1 are structurally homogeneous contiguous portions of the dorsomedial cortex. As this is reorganized into the mammalian hippocampus, CA3 and CA1 differentiate in two important ways. First, only CA3 receives the projections from the dentate gyrus, the mossy fibers. Second, only CA3 is dominated by recurrent collaterals, whereas most of the inputs to CA1 cells are the projections from CA3, the Schaffer collaterals (Amaral et al. 1990).

The simplest version of the postprocessor notion is that it may be useful to add a further feedforward associative network, to clean up memory representations already retrieved, but in incomplete form, by the CA3 network. The extra stage of recoding, if based on more neurons (there are more pyramidal cells in CA1 than in CA3 across all species where numbers have been estimated) could also add robustness to the retrieved representation. Yet a mathematical network analysis of the cleanup notion—in the framework of discrete "episodic memory" fixed-point attractors and neglecting the separate entorhinal cortex inputs directly to CA1—failed to illustrate impressive advantages to adding such a post-processing stage (Treves 1995). Information content grows from CA3 to CA1, but by a minor amount.

A more interesting suggestion comes from a review of neuropsychological studies in rats (Kesner et al. 2002) that indicate a more salient role for CA1 along the temporal dimension. CA3 may specialize in associating information that was experienced strictly at the same time, whereas CA1 may more than CA3 link together information across adjacent times. This may lead to the storage of sequences of instantaneous events, that together build up an episode, or, if the events are not parsed, to effectively continuous attractors along the temporal dimension. A way to formulate a qualitative implication of such a putative functional differentiation is to state that CA1 is important for prediction, i.e., for producing an output representation of what happened just after whatever is represented by the pattern of activity retrieved at the CA3 stage. Note, however, that reading the review by Kesner and colleagues (2002) in full indicates that the table at the end is a well-meaning simplification. Their figure 31.2 suggests that CA3 may be involved in temporal pattern separation just as much as CA1. Moreover, the role of either DG or CA3 in temporal pattern association has not been satisfactorily assessed. Further, available studies on the role of CA1 fail to make a clear distinction between tasks in which massive hippocampal outputs to the cortex are crucial and tasks in which a more limited hippocampal influence on the cortex may be sufficient. In the first case, lesioning CA1 should have an effect independent of what CA1 specifically contributes to information processing, simply because one is severing the main hippocampo-cortical output pathway. In the second, CA3 outputs through the fimbria/fornix could enable hippocampus-mediated influences to be felt, even if the specific CA1 contribution is absent.

Testing the Prediction of Predictive Coding

I have explored the hypothesis that the differentiation between CA3 and CA1 may help solve precisely the computational conflict between pattern completion, or integrating current sensory information on the basis of memory, and prediction, or moving from one pattern to the next in a stored continuous sequence. To obtain results comparable with typical rat experiments, I have used the same neural network simulations of a virtual rat exploring a small toroidal environment as the ones analyzing the role of dentate inputs to CA3 (Treves 2004). The network model was thus trained to acquire a chart representation of the explored environment as a spatially continuous attractor. Temporal continuity along each trajectory was used to assess the extent to which CA3 would take care of (spatial) pattern completion, while CA1 would concentrate on prediction (i.e., temporal pattern completion). With the simulations one can, at the price of some necessary simplification, compare the performance of the differentiated circuit with a "uniform," nondifferentiated circuit of equal number and type of components (one in which CA3 and CA1 have identical properties, e.g., both receive mossy fibers and are interconnected with recurrent collaterals). Lesion studies, by contrast, can only compare the normal circuit with others with missing components, making it difficult to assess the significance of a differentiation.

The functional differentiation hypothesis was not really convincingly supported by neural network simulations. The conflict between spatial pattern completion, as quantified by localization accuracy, and temporal prediction indeed exists, but two mechanisms that would more directly relate to a functional CA3-CA1 differentiation were found unable to produce genuine prediction. Instead, a simple mechanism based on firing-frequency adaptation in pyramidal cells was found to be sufficient for prediction, with the degree of adaptation as the crucial parameter balancing retrieval with prediction. This is evident from the simulations of the nondifferentiated model. The differentiation between the connectivity of CA3 and CA1 does not really influence the predictiveness, or degree of anticipation, of hippocampal activity. The differentiation has a significant positive effect, however, and, in particular for a given anticipatory interval, it significantly increases, in the model, the information content of hippocampal outputs, making the CA1 representation more informative than the CA3 one (or the nondifferentiated one) when used to decode the position of the virtual rat. Different degrees of adaptation in CA3 and CA1 cells were not, however, found to lead to better performance, further undermining the notion of a full qualitative functional dissociation. There may, therefore, be just a plain quantitative advantage in differentiating the connectivity of the two fields, just as the hypothesis about isocortical lamination holds that there may be just a plain quantitative advantage in differentiating connectivity across layers. In a sense, the outcome of the simulations supports a revised version of the postprocessing cleanup notion. As figure 3.4 shows, the information content in CA1 in the differentiated model is higher than in CA3, with the nondifferentiated model midway between the two.

Correlated Environments Stimulate Orthogonal Ideas

As for the lamination study, the analysis of this hypothesis about the differentiation of hippocampal subfields was based on the simulation of two simplified models, uniform and differentiated, tested on the same task, in this case acquiring a memory chart for a single spatial environment. The accuracy of spatial memory retrieval is subject to the general "spin-glass" limit, and it is further modulated by connectivity details. Recent results obtained recording the activity of multiple hippocampal cells in the labs of Edvard and May-Britt Moser (Leutgeb et al. 2004) and of James Knierim (Lee et al. 2004) indicate a

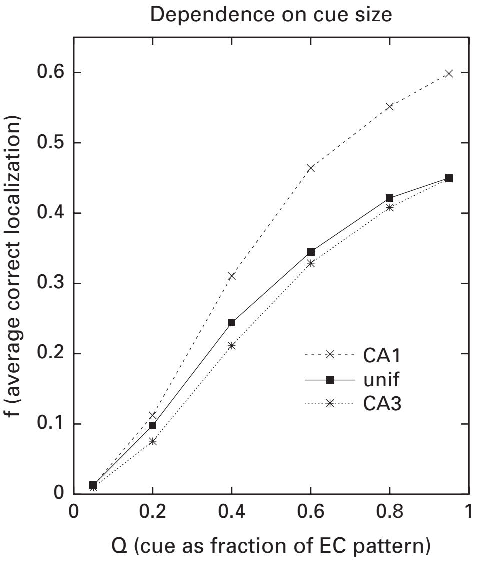

Figure 3.4
Localization accuracy as a function of cue size, after extensive training, in hippocampal network models with and without the differentiation between CA3 and CA1 (and with mossy fiber inputs). In the differentiated model, the position of the virtual rat is better decoded from the CA1 than from the CA3 representation. Simulation details in Treves (2004).

potentially much more dramatic differentiation between CA3 and CA1 units, which has to do with their ability to distinguish among several spatial environments. Activity in CA3 and CA1 was found to differ remarkably when rats were asked to explore environments that some cues suggested were the same, and others, that they were different. CA3 appears to take an all-or-none decision, usually allocating nearly orthogonal neural representations to even very similar environments and switching to essentially identical representations only above a high threshold of physical similarity. Activity in CA1, on the other hand, varies smoothly to reflect the degree of similarity. This functional differentiation and the finding that new representations in CA3 emerge slowly, presumably through iterative processing, are entirely consistent with the recurrent character of the CA3 network and the prevailing feedforward character of the CA1 network. Further surprises have come from applying a "morphing" paradigm, to test spatial representations in environments quasi-continuously changed between two well-learned extremes (Willis et al. 2005; Leutgeb et al. 2005).

In their original form (Treves 2004), the connectivity differentiation models addressed the mechanism linking firing-rate adaptation to the prediction of spatial position within a single environment but could not capture any advantage brought about by the connectivity differentiation having to do with multiple maps. The experimental results have stimulated the development of more elaborate computational models, which however still have to satisfactorily find their way around the spin-glass limit on memory retrieval (see Papp et al. 2007). In fact, training virtual rats on several virtual environments, correlated or not, requires them to be endowed with large virtual brains. Simulations with networks of a thousand units or so, which were adequate for the single-environment case, have to be extended to networks larger by one or two orders of magnitude, which have become time-consuming to simulate extensively. Even then, because of the heavy memory load for multiple environments (Battaglia and Treves 1998), the representations tend to collapse on each other, making the comparison with real rat data more problematic (Papp and Treves 2008). One observation that emerges from this study, already at this stage, is that CA3 representations tend to be more fragmented, in the sense that neural activity in pairs of separate locations can be identical, or quite different, in violation of the metric nature of the environment. CA1 representations tend to be relatively smoother, with a higher match between the distance among locations and the difference among their neural activity vectors. This smoothing function for CA1 strongly resembles the cleanup notion originally investigated for discrete memories, which was relevant also to the notion of prediction, that implies continuity in time but which now finds a more interesting role in reproducing the continuity of physical space. Thanks to the experimental findings with correlated environments, we may be beginning to finally "understand" CA1 and to make some (spatial) sense of the events that drastically altered the structure of our medial pallium hundreds of millions of years ago.

Quality vs. Quantity and the Need to Adapt

All three studies reviewed here require firing-rate adaptation as a crucial ingredient in producing, respectively, a separation between the processing of "what" and "where" information, transitions to different semantic attractor states, and the prediction of future locations in a spatial environment. In all three, memory retrieval is limited by the "spin-glass" constraint. A fundamental dissimilarity is in the relation between qualitative and quantitative changes. In the two "mammalian" studies, the hypothesis is that a major *qualitative* structural change may have served to produce a solely *quantitative* functional advantage. Although the first such hypothesis seems a posteriori more convincing than the second, both are methodologically valid a priori, and in fact it has been noted (Carroll 1988) that often in evolution major steps may subserve only "small" improvements in survival ability. In the "human" study, the hypothesis considered has the opposite flavor: a quantitative change in connectivity (admittedly, a *major* change) would be enough to produce a phase transition to an entirely novel computational faculty—namely, infinite recursion—with its collateral effects including the emergence of language in humans. Although all these hypotheses require much further testing, they serve to underscore the often subtle relations between structure and function that can apply to cortical networks, mediated by the collective emergent dynamics of large populations of neurons.

Acknowledgments

I am grateful to colleagues in the lab of Edvard and May-Britt Moser for letting me participate in their investigations.

References

Abeles M (1991) Corticonics: Neural Circuits of the Cerebral Cortex. Cambridge: Cambridge University Press.

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

Amit DJ (1989) Modeling Brain Function. Cambridge: Cambridge University Press.

Amit DJ, Gutfreund H, Sompolinsky H (1987) Statistical mechanics of neural networks near saturation. Ann Phys (N.Y.) 173: 30–67.

Barton RA (2007) Evolutionary specialization in mammalian cortical structure. J Evolution Biol 20: 1504–1511.

Battaglia FP, Treves A (1998) Attractor neural networks storing multiple space representations: A model for hippocampal place fields. Phys Rev E 58: 7738–7753.

Bingman VP, Jones T-J (1994) Sun-compass based spatial learning impaired in homing pigeons with hippocampal lesions. J Neurosci 14: 6687–6694.

Bingman VP, Sharp PE (2006) Neuronal implementation of hippocampal-mediated spatial behavior: A comparative evolutionary perspective. Behav Cogn Neurosci Rev 5: 80–90.

Bininda-Emonds ORP, Cardillo M, Jones KE, MacPhee RDE, Beck RMD, Greyner R, Price SA, Vos RA, Gittelman JL, Purvis A (2007) The delayed rise of present-day mammals. Nature 446: 507–512.

Bliss TV, Lømo T (1973) Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. J Physiol 232: 331–356.

Braitenberg V, Schüz A (1991) Anatomy of the cortex. Berlin: Springer-Verlag.

Carroll RL (1988) Vertebrate paleontology and evolution. New York: WH Freeman.

Clayton NS, Bussey TJ, Dickinson A (2003) Can animals recall the past and plan for the future? Nat Rev Neurosci 4: 685–691.

Clayton NS, Krebs JR (1995) Memory in food-storing birds: From behaviour to brain. Curr Opin Neurobiol 5: 149–154.

Collingridge GL, Bliss TV (1995) Memories of NMDA receptors and LTP. Trends Neurosci 18: 54–56.

Diamond IT, Conley M, Itoh K, Fitzpatrick D (1985) Laminar organization of geniculocortical projections in *Galago senegalensis* and *Aotus trivirgatus*. J Comp Neurol 242: 584–610.

Diamond IT, Hall WC (1969) Evolution of neocortex. Science 164: 251–262.

Eichenbaum H (2000) A cortical-hippocampal system for declarative memory. Nat Rev Neurosci 1: 41–50.

Elston GN (2000) Pyramidal cells of the frontal lobe: All the more spinous to think with. J Neurosci 20: RC95(1–4).

Finlay BL, Darlington RB (1995) Linked regularities in the development and evolution of mammalian brains. Science 1268: 1578–1584.

Fyhn M, Hafting T, Treves A, Moser EI, Moser M-B (2007) Hippocampal remapping and grid realignment in entorhinal cortex. Nature 446: 190–194.

Fyhn M, Molden S, Witter MP, Moser EI, Moser M-B (2004) Spatial representation in the entorhinal cortex. Science 305: 1258.

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

Goffinet AM (2006) What makes us human? A biased view from the perspective of comparative embryology and mouse genetics. J Biomed Discov Collab 1: 16.

Hafting T, Fyhn M, Molden S, Moser M-B, Moser EI (2005) Microstructure of a universal spatial map in the entorhinal cortex. Nature 436: 801–805.

Hasselmo ME, Schnell E (1994) Laminar selectivity of the cholinergic suppression of synaptic transmission in rat hippocampal region CA1: Computational modeling and brain slice physiology. J Neurosci 14: 3898–3914.

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

Hasselmo ME, Wyble B, Wallenstein G (1996) Encoding and retrieval of episodic memories: Role of cholinergic and GABAergic modulation in hippocampus. Hippocampus 6: 693–708.

Hauser MD, Chomsky N, Fitch WT (2002) The faculty of language: What is it, who has it, and how did it evolve? Science 298: 1569–1579.

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

Jerison HJ (1990) Fossil evidence of the evolution of the brain. In: Comparative structure and evolution of cerebral cortex, vol. 8A (Jones EG, Peters A, eds) 285–309. New York: Plenum Press.

Kanter I (1988) Potts-glass models of neural networks. Phys Rev A 37: 2739–2742.

Kesner RP, Gilbert PE, Lee I (2002) Subregional analysis of hippocampal function in the rat. In: Neuropsychology of memory (Squire LR, Schacter DL, eds), 395–411. New York: Guilford Press.

Kropff E, Treves A (2005) The storage capacity of Potts models for semantic memory retrieval. J Stat Mech 2: P08010.

Krubitzer L, Huffman KJ (2000) Arealization of the neocortex in mammals: Genetic and epigenetic contributions to the phenotype. Brain Behav Evol 55: 322–335.

Laplagne DA, Esposito MS, Piatti VC, Morgenstern NA, Zhao C, van Praag H, Gage FH, Schinder AF (2006) Functional convergence of neurons generated in the developing and adult hippocampus. PLoS Biol 4: e409.

Lassalle J-M, Bataille T, Halley H (2000) Reversible inactivation of the hippocampal mossy fiber synapses in mice impairs spatial learning, but neither consolidation nor memory retrieval, in the Morris navigation task. Neurobiol Learn Mem 73: 243–257.

Lee I, Kesner RP (2004) Encoding versus retrieval of spatial memory: Double dissociation between the dentate gyrus and the perforant path inputs into CA3 in the dorsal hippocampus. Hippocampus 14: 66–76.

Lee I, Yoganarasimha D, Rao G, Knierim JJ (2004) Comparison of population coherence of place cells in hippocampal subfields CA1 and CA3. Nature 430: 456–459.

Leutgeb JK, Leutgeb S, Moser M-B, Moser EI (2007) Pattern separation in the dentate gyrus and CA3 of the hippocampus. Science 315: 961–966.

Leutgeb S, Leutgeb JK, Treves A, Moser M-B, Moser EI (2004) Distinct ensemble codes in hippocampal areas CA3 and CA1. Science 305: 1295–1298.

Leutgeb JK, Leutgeb S, Treves A, Meyer R, Barnes CA, McNaughton BL, Moser M-B, Moser EI (2005) Progressive transformation of hippocampal neuronal representations in "morphed" environments. Neuron 48: 345–358.

Mallamaci A, Stoykova A (2006) Gene networks controlling early cerebral cortex arealization. Eur J Neurosci 23: 847–856.

Marr D (1971) Simple memory: A theory for archicortex. Philos T Roy Soc Lond B 262: 24–81.

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

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

Papp G, Treves A (2008) Network analysis of the significance of hippocampal subfields. In: Hippocampal place-fields: Relevance to learning and memory (Mizumori S, ed). Oxford: Oxford University Press.

Papp G, Witter MP, Treves A (2007) The CA3 network as a memory store for spatial representations. Learn Memory 14: 732–744.

Ramirez-Amaya V, Marrone DF, Gage FH, Worley PF, Barnes CA (2006) Integration of new neurons into functional neural networks. J Neurosci 26: 12237–12241.

Rolls ET (1989) Functions of neuronal networks in the hippocampus and cerebral cortex in memory. In: Models of brain function (Cotterill R, ed), 15–33. New York: Cambridge University Press.

Roudi Y, Treves A (2006) Localized activity profiles and storage capacity of rate-based autoassociative networks. Phys Rev E 73: 061904.

Samsonovich A, McNaughton BL (1997) Path integration and cognitive mapping in a continuous attractor neural network model. J Neurosci 17: 5900–5920.

Skaggs WE, Knierim JJ, Kudrimoti HS, McNaughton BL (1995) A model of the neural basis of the rat's sense of direction. In: Advances in neural information processing systems (Tesauro G, Touretzky D and Leen T, eds), 173–180. Cambridge, MA: MIT Press.

Sompolinsky H, Shapley R (1997) New perspectives on the mechanisms for orientation selectivity. Curr Opin Neurobiol 7: 514–522.

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

Treves A (2003) Computational constraints that may have favoured the lamination of sensory cortex. J Comput Neurosci 14: 271–282.

Treves A (2004) Computational constraints between retrieving the past and predicting the future, and the CA3-CA1 differentiation. Hippocampus 14: 539.

Treves A (2005) Frontal latching networks: A possible neural basis for infinite recursion. Cognitive Neuropsych 21: 276–291.

Treves A, Cerasti E, Papp G (2008) The dentate gyrus and the formation of new spatial representations in CA3. 6th FENS Forum abstract 225.25.

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 4: 374–391.

Treves A, Tashiro A, Witter ME, Moser EI (2008) What is the mammalian dentate gyrus good for? Neuroscience, Forefront Review, in press.

Ulinski PS (1990) The cerebral cortex of reptiles. In: Cerebral cortex, volume 8A: Comparative structure and evolution of cerebral cortex (Jones EG, Peters A, eds), 139–215. New York: Plenum Press.

Ullman MT (2001) A neurocognitive perspective on language: The declarative/procedural model. Nat Rev Neurosci 2: 717–726.

Willis TJ, Lever C, Cacucci F, Burgess N, O'Keefe J (2005) Attractor dynamics in the hippocampal representation of the local environment. Science 308: 873–876.

Willshaw D, Buckingham J (1990) An assessment of Marr's theory of the hippocampus as a temporary memory store. Philos T Roy Soc Lond B 329: 205–215.

Wilson EO (1975) Sociobiology. The new synthesis. Cambridge, MA: Harvard University Press.

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

Yamamori T, Rockland KS (2006) Neocortical areas, layers, connections, and gene expression. Neurosci Res 55: 11–27.