CA3 and Memory/Review

The CA3 network as a memory store for spatial representations

Gergely Papp, 1 Menno P Witter, 2,3 and Alessandro Treves 1,2,4

1 Scuola Internazionale Superiore di Studi Avanzati (SISSA), Cognitive Neuroscience Sector, Trieste 34014, Italy; 2 Norwegian University of Science and Technology (NTNU), Kavli Institute for Systems Neuroscience and Centre for the Biology of Memory, Trondheim NO-7489, Norway; 3 VU University Medical Center (VUMC), Department of Anatomy & Neuroscience, Amsterdam 1081 BT, The Netherlands

Comparative neuroanatomy suggests that the CA3 region of the mammalian hippocampus is directly homologous with the medio-dorsal pallium in birds and reptiles, with which it largely shares the basic organization of primitive cortex. Autoassociative memory models, which are generically applicable to cortical networks, then help assess how well CA3 may process information and what the crucial hurdles are that it may face. The analysis of such models points at spatial memories as posing a special challenge, both in terms of the attractor dynamics they can induce and how they may be established. Addressing such a challenge may have favored the evolution of elements of hippocampal organization observed only in mammals.

Introduction: A comparative perspective

The forebrain of vertebrates shows remarkable morphological variation and specialized adaptations, presumably related to differences in environment. Yet, in all species part of the brain apparently is involved in the formation of map-like representations of the environment. The need to localize distributed resources such as food, shelter, and so on is an essential component of the evolutionary success of freely moving species. It is quite likely that the formation of map-like representations strongly depend on the potential to generate relational representations of environmental features (of course, in order to navigate, animals do not simply rely on these relational maps between landmarks, but also make use of, e.g., idiothetic cues).

In all species studied, including teleost fish, amphibians, reptiles, and mammals, relational representations are mediated by structures in the brain that embryologically derive from the most medial part of the telencephalic anlage (Rodriguez et al. 2002). In mammalian species, this derivative is called the hippocampal formation, and it is of interest that a correlation apparently exists between the volume of the hippocampal formation and successful spatial performance. It is therefore not surprising that the mammalian hippocampal formation is considered to represent one of the phylogenetically oldest cortical areas. Comparative neuroanatomical studies, in conjunction with embryological studies, as well as data on gene-expression patterns have indicated that the hippocampal formation in mammals most likely is homologous to parts of the medio-dorsal cortex in reptiles (Lopez-Garcia and Martinez-Guijaro 1988; Ulinsky 1990a,b; ten Donkelaar 2000). The hippocampal formation originates from the medio-dorsally positioned anlage in the developing brain, which shows a gross overall similarity with the dorso medial cortex in reptiles (Stephan 1975). In birds, the cortex of the medial surface of the pallium merges with more ventrally located pallial structures, such that the apparent typical layered structure of the cortex is no longer apparent.

The overall wiring of the mammalian hippocampal formation (Amaral and Witter 1989) will not be reviewed here. Suffice it to note that, both with respect to its extrinsic as well as its intrinsic connections, it shows remarkably strong similarities even when such different species as the mouse and monkey are compared. This indicates that there may be some ecological benefit for this stabilized connectivity. In order to gain insight into the functional relevance of this particular organization of the hippocampal system, only a few approaches are at hand. Either we test the effects of loss of parts of the system, for example, in lesioned animals (Kesner et al. 2002), or we adopt computational modeling approaches, in which, at the cost of many necessary simplifications, network organization can be varied at will. Both approaches have to be informed by an appreciation of the organization of homologous structures in nonmammalian species of vertebrates that indicates which changes and refinements may have been driven by evolutionary pressures.

The medial pallium in reptiles and in birds

Structure and connectivity of the mediadorsal cortex of reptiles

The cortex in reptiles is generally divided into mediadorsal cortex, dorsal, and lateral cortex. The mediadorsal cortex is further subdivided into a small-celled part and a large-celled part, also referred to as medial and mediadorsal cortex (Fig. 1A,B). Like the cortex that makes up the mammalian hippocampus, within the reptilian cortex three different layers are distinguished, an outer molecular layer, the cell layer, and, bordering the underlying white matter, the polymorph layer (Fig. 1D). In general, the small-celled portion of the reptilian cortex is considered to be comparable to the dentate gyrus of the mammilian hippocampus. The large-celled part is taken to be comparable to the CA fields, but in reptiles, no further subdivisions have been distinguished, for example, between a CA3-like and a CA1-like region.

Neurons in the small celled portion are pyramidal or spherical, with an average diameter of ~6–18 µm. They are closely packed and extend dendrites into the molecular layer as well as into the polymorph layer. Similar to what is seen in the dentate gyrus in mammals, a majority of the dendrites do extend into the molecular layer, and the first bifurcation is close to the soma. In contrast to the mammalian situation though, the cell types appear to be more variable, such that a total of six different cell types have been described within the cell layer, of which only

*Corresponding author.
E-mail ale@sissa.it; fax +39-040-3787615.
Article is online at http://www.learnmem.org/cgi/doi/10.1101/lm.687407.
some contribute to projections targeting adjacent portions of the
cortex, mainly to the large-celled part of the mediodorsal cortex
(Wouterlood 1981; Olucha et al. 1988). These projections are
strongly Timm positive, strikingly similar both morphologically
as well as in terms of overall distribution to the mossy fiber pro-
jection in mammals (Lopes Garcia and Martinez-Guijaro 1988;
Olucha et al. 1988). Although not much is known about the
different local circuitry neurons in the reptilian brain, neurons
have been described in both the molecular layer and the deep
polymorph layer.

The cell layer of the large-celled portion comprises mainly
one cell type, which shows a polygonal or pyramidal shape with
large apical dendrites extending into the molecular layer as well
as basal dendrites extending into the polymorph layer and adja-
cent white matter. Short basal dendrites extend from the soma
parallel to the cell layer. Additional neurons in the molecular
and polymorph layer have been described as well (Wouterlood 1981;
ten Donkelaar 2000). Note that in several mammalian species,
the anterior (supracallosal) continuation of the hippocampus,
dindusium griseum, and tenia tecta (considered the olfactory hip-
locampus) shows similarities to the lizard medial cortex, where
the dentate and CA fields form a continuous sheet of cells with
two morphologies, granule and pyramidal (Stephan 1975; Wyss
and Stripanidkulchai 1983; Shipley and Adamek 1984; Künzle
2004). Projections distribute widely in the cortex, including a
return projection to the small-celled part. Most notably, wide-
spread dense intrinsic projections have been described as distrib-
uting both to apical as well as basal dendrites, reminiscent of the
associative connectivity in the mammalian CA3 (Olucha et al.
1988; Hoogland and Vermeulen-Vanderzee 1993).

Structure and connectivity of the dorsomedial
telencephalon of birds
In birds, the dorsomedial telencephalon is considered the homo-
log of the hippocampal region of mammals. However, a detailed
comparison between subdivisions of the avian and mammalian
hippocampus, and even between birds and reptiles, is still specu-
lative. According to the detailed descriptions of Ariens-Kapper et
al. (1936), there is a layered cortical structure that might be con-
considered the avian hippocampus and parahippocampal region.
In the chicken, the dorsomedial cortex comprises a superficial plexi-
form layer, a granular layer, and a periventricular or polymorph
layer (Molla et al. 1986). Interestingly, according to the latter
authors, the cells in the granular layer are pyramidal or bipyr-
maidial cells, similar to what has been reported for the mediodorsal
cortex in a number of reptilian species. On the basis of afferent
and efferent connections, it can also be argued that the dorso-
medial telencephalic area in birds is comparable to the hippo-
campal formation and parahippocampal region in mammals (for
review, see Dubbeldam 1998).

Two issues appear relevant here. First, the question arises
whether or not in birds hippocampal and parahippocampal sub-
divisions are apparent that show cytoarchitectonic and/or con-
nectional similarities to those described in mammals; second, it
is of interest to assess whether the typical unidirectional connec-
tivity seen in the mammalian hippocampal system is present in
birds. Unfortunately, staining the avian brain for the distribution
of Zinc with the Timm stain, which in mammals and reptiles
stains the projection from neurons in the granular cell layer to a
cluster of large pyramidal cells, does not provide useful clues. In
birds the Timm stain only results in rather diffuse and weak stain-
ing (Faber et al. 1989; Montagnese et al. 1993, 1996).

A case of mistaken identity?
Two recent studies addressed both issues by studying the con-
nectional organization of the dorsomedial telencephalic domain
of the pigeon in much detail (Kahn et al. 2003; Atoji and Wild
2004). Both studies describe, within this part of the avian brain,
A V-shaped ventral medial region subdivided into a lateral and
medial blade of neurons that enclose an area called central or
triangular region (cf. Fig. 1C). This ventral region is bordered
more dorsally by an area referred to as dorsomedial cortex, fol-
lowed by a more dorsolaterally positioned dorsolateral cortex.
Both follow the original descriptions of Karten and Hodos (1967),
supplemented with neurochemical patterns as described by
Erichsen et al. (1991). On the basis of the connectional data, both
studies describe a series of connections comparable to the mam-
malian trisynaptic circuit. Unfortunately, the results lead Kahn
et al. (2003) to suggest that the ventral domain is comparable to
CA1, similar to suggestions based on anterograde tracing in the
zebra finch (Székely and Krebs 1996), whereas Atoji and Wild
(2004) suggest that this region actually represents the dentate
 gyrus. In terms of position and shape, the suggestion that the
ventral region actually is the “dentate” of the avian brain is the
more apparent one (see also Atoji and Wild 2006); however, no
clear conclusions can be stated here at this point in time. More-
over, both in birds and in reptiles, the dorsomedial cortex issues
projections to the lateral septum, similar to what is reported for
CA3 and CA1 in mammals. Irrespective of whether the ventral
of dorsomedial areas should be considered the homolog of the
mammalian dentate gyrus, neither of the connectional studies
indicate that in the avian brain unidirectional connections of
one cortical area to another are present that form large, Zinc-
positive terminals that may function as detonator synapses, as
proposed for the synapses on the mossy fibers to CA3 cells in
mammals (Andersen and Loyning 1962). In all reports, bidirec-
tional connections are, in fact, predominant.

Physiology and plasticity of the avian
hippocampal formation
In comparison with the vast amount of neurophysiological data
in mammals, there are relatively few studies analyzing the activ-
ity of neurons in the avian hippocampal formation, and they

www.learnmem.org 733 Learning & Memory
point at broad similarities with the spatially selective activity of mammalian hippocampal neurons. These similarities are consistent with the intensely studied role of the avian hippocampus in episodic and spatial memory (Krebs et al. 1989; see the discussion by Clayton et al. 2003; Healy et al. 2007). As reviewed by Bingman and Sharp (2006), location-specific cells, “arena-off” cells, and path cells have been described in the pigeon hippocampus and even some putative “grid-like” cells. Recordings have also revealed the presence of oscillatory activity in birds, similar to the theta rhythm of rodents, though slower (Siegel et al. 2000). Different cell types, however, could not be reliably localized to distinct subdivisions within the hippocampus (see Hough and Bingman 2004) or to a distinct layer within a subdivision, with only some distinction based on electrophysiological cell properties, perhaps corresponding to principal cells and interneurons (Siegel et al. 2002). Units in the ventral regions might tend to be similar in spike width and frequency to the pyramidal cells of CA regions in the rat, while units in the dorsocaudal region might be more similar, in terms of firing pattern, to the granule cells of the mammalian DG, perhaps in support of the conclusions of Kahn et al. (2003). If anything, there is clearer evidence for lateralization, e.g., path cells are found on the left side, and location-specific cells show more prominent selectivity on the left (Siegel et al. 2006).

Forms of synaptic plasticity potentially involved in learning and memory such as LTP and LTD have also been discovered in the avian brain (Scott and Bennett 1993; Wierszko and Ball 1993; Margrie et al. 1998). There is also evidence for the formation of new synapses after training (Únal et al. 2002) as well as considerable long-standing evidence for neurogenesis, which antedates parallel evidence more recently found in mammals. Nevertheless, in contrast to the extensive literature available for rodents, synaptic plasticity has not been described in subregional detail.

Homology and beyond
In conclusion, our current understanding does not allow us to draw firm conclusions concerning structures in the avian brain that would represent the mammalian dentate gyrus, despite the homology between the overall “hippocampal region” in the two lineages (Colombo and Broadbent 2000). Importantly, some aspects of the connectivity are quite different, leading one to doubt the relevance of forcing a correspondence at the subregional level, given hundreds of million years of separate evolution (Striedter 2005). The salient exception is that, as for the reptilian, the avian hippocampus appears to retain a connectional system comparable to the autoassociative system of the mammalian CA3 field (Fig. 1E). This suggests that a fundamental associative memory function of the primitive cortex (Braitenberg and Schütz 1991) may have been largely preserved in all amniotes, based on its recurrent architecture and associative synaptic plasticity, though perhaps made more complex and powerful in different ways in reptilian, avian, and mammalian derivatives. This simplifying perspective, which regards the CA3 field as part of the common amniotic heritage, and DG and CA1 as novel mammalian concoctions (although with similarities, particularly in the distribution of Zinc, with reptiles), informs the computational analyses described next.

The CA3 region as an autoassociative memory
David Marr brilliantly synthesized ideas about the role of the hippocampus in memory formation, which he may have subsumed indirectly from neuropsychological studies, and took them as the starting point to understand the organization of hippocampal circuits (Marr 1971). This “structure-from-function” theoretical research program has been enormously influential, even though the details of his modeling approach are difficult to appraise. For example, Marr eloquently emphasized, in words, the “collateral effect,” i.e., the potential role in pattern completion of recurrent connections, prominent among CA3 pyramidal cells (Amaral et al. 1990); but, his own model was not really affected by the presence of such collaterals, as shown later by careful meta-analysis (Willshaw and Buckingham 1990). Marr thought in terms of discrete memory states, and devoted an entire section of his work to “capacity calculations,” which indicates that he realized how an estimate of the maximum number of activity patterns retrievable from a memory network could be a central contribution of mathematical models. The simple model he considered, with binary processing units and binary synaptic weights, is endowed with the capacity to retrieve up to \( p_r \) patterns, where \( p_r (a_{CA})^2 \ll 1 \) and \( a_{CA} \) is the activity level of CA pyramidal units, i.e., the fraction of units active in a pattern (which generalizes to the sparsity parameter for nonbinary patterns). Marr stated that with an activity level \( a_{CA} = 0.001 \), the hippocampus could store and retrieve of the order of \( p_r = 100,000 \) memories, which he reckoned was a reasonable number, at least for a temporary memory store. Those “doubly binary” models cannot be easily related to real CA3 neuronal networks, however; even if they were, the observed mean sparsity \( a_{CA} = 0.03-0.04 \) (see, e.g., Papp and Treves 2007) would lead to a rather dismal capacity of about \( p_r = 100 \) patterns. Although the work by Marr was nearly simultaneous with the discovery of place cells (O’Keefe and Dostrovsky 1971) and with that of long-term synaptic potentiation (Bliss and Lomo 1973; cited as a note added in proof), for a long time Marr 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).

The dentate gyrus as a generator of CA3 activity
With their 1987 review, McNaughton and Morris (1987) re-examined and revived the Marr framework, discussing several “Hebb-Marr” associative-memory model architectures and whether they resembled hippocampal networks. The operation of such models of attractor networks (Amit 1989) can be more readily analyzed if the memory patterns to be stored are assigned “by hand,” rather than self-organized under the influence of ongoing inputs. One can imagine that a system of one-to-one connections from another area may effectively “transfer” a pattern of activity from the other area, where it is determined by some unspecified process, to the associative memory network. McNaughton and Morris (1987) took the strong “detonator” synapses on the MF projections from DG to CA3 (Andersen and Loyning 1962) as an approximate implementation in the real brain of such one-to-one connections. The distributions of activity to be stored in memory would be effectively generated in DG, perhaps by expansion encoding (as hypothesized for granule cells in the cerebellum) and then simply transferred to CA3.

As clarified by Treves and Rolls (1992), for the dentate gyrus to “impose” a novel pattern of activity onto CA3, it need not transfer its own and one-to-one connections are not necessary: they would not be very explanatory either, in that they would simply anticipate the question of how to generate an appropriate pattern of activity back to DG, rather than resolving it. Instead, what matters is that the MF synapses be strong, sparse, and conveying sparse activity from DG. This is sufficient to effectively select a limited ensemble of CA3 cells to represent a new memory unrelated to ensembles that are coactivated in other, previously stored memories, and which would tend to be reinstated by the collateral effect. The competition between the MF projections,
forcing a novel ensemble, and the recurrent connections, reinstating fragments of previously stored ones, is a quantitative balance, which has to be shifted toward the MF inputs only when the system is in “storage mode.”

When retrieving a previously stored memory representation, the input to the recurrent network has to relay a cue, expressed as a pattern of activity on the afferent fibers correlated with the pattern at the time of storage of a particular memory. Simple algebra shows that such correlation tends to be washed away unless the afferents make many associatively modifiable synapses onto each receiving cell, synapses that have to be modified during storage (Treves and Rolls, 1992). The CA3 region would then work better as a recurrent associative memory, implementing pattern completion thanks to the collateral effect (Rolls 1989) if it had two separate afferent systems: one particularly strong at the time of storage, relaying sparse activity, and possibly weak at retrieval—which can be identified with the mossy fibers—and one relayed by many synapses, highly plastic but weak during storage, and transmitting at retrieval, to be identified with the perforant path.

The argument requires DG activity to be sparse or moderately sparse (values of \(a_{DG} = 0.1\) or 0.02 produce nearly identical results in the mathematical model), but it does not constrain further the form of granule cell activity, e.g., in its spatial correlates. It does predict, however, that since the useful role of the “duplicated” DG input is only in establishing new CA3 representations, lesioning DG or blocking MF transmission should have no effect on memory retrieval. The prediction is so far consistent with behavioral results obtained in two independent experimental approaches (Lassalle et al. 2000; Lee and Kesner 2004).

The function of neuromodulation

A selective modulation of the activity (and plasticity) of specific synaptic systems may be effected by acetylcholine (ACh), exploiting the orderly arrangement of pyramidal cell dendrites in the cortex, which allows for differential action on the synapses distributed in distinct layers (Haselgrove and Schnell 1994). Acetylcholine is one of several very ancient neuromodulating systems, well conserved across vertebrates, and may have operated in this way already in the early reptilian cortex, throughout its subdivisions. Hasselmo has emphasized this likely 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 the 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. A drawback of relying on ACh modulation alone is that it requires an active process that distinguishes storage from retrieval periods, and regulates ACh-release accordingly. Combining ACh modulation with MF orthogonalization may allow the hippocampus to operate efficiently also during behaviors, such as exploration, when storage and retrieval are perhaps largely admixed. In addition, the more complex circuitry allows for further neuromodulatory control, e.g., through dopamine release (Kobayashi and Suzuki 2006). In refining the architecture of the hippocampus, therefore, it could be that mammals have devised a more efficient memory process, which augments rather than replaces the earlier one based on neuromodulation.

Quantitative analyses of storage capacity

Such a hypothesis leads to predict quantitative rather than qualitative effects for most manipulations, and, hence, it has to be articulated in terms of network models that allow quantitative measures. Unfortunately, the rudimentary binary-synapse model considered by Marr (1971) does not allow a meaningful correspondence with experimental measures, as indicated by the many predictions at the end of his work, which have remained untested. The associative memory model introduced by Hopfield (1982), on the other hand, proved suitable for a theoretical breakthrough when it was analyzed by Amit et al. (1987). They were able to calculate its storage capacity by applying mathematical techniques from statistical physics, which could later be applied to biologically more plausible versions of the model. The generic result is that a network with \(C\) recurrent connections per unit can retrieve, i.e., complete, up to

\[
p_c = 0.2-0.3 \frac{C}{a} \ln (1/a),
\]

patterns of activity, where \(a\) is their sparsity value (in the sparse, \(a < 1\) regime) (see Treves and Rolls 1991). Each such representation can contain at least

\[
i = N a \ln (1/a) \text{ bits}
\]

of new information, where \(N\) is the number of units. This leads to a maximum total amount of information, cumulating the contribution of all memory patterns, which is proportional to the number of synapses \(NC\), and which does not depend much on sparsity

\[
I_{\max} = 0.2-0.3 NC \text{ bits}.
\]

Efficient use of the retrieval capacity of the CA3 recurrent network requires that its pyramidal units encode as much new information in a pattern of activity as the amount estimated in Equation 2, which can be retrieved later, just as efficient use of cash-lending machines requires one to deposit in the corresponding account the wherewithal that one wants to withdraw later. The challenge for afferent inputs is then to prevail during storage over the recurrent connections, which do not impart new contents to a pattern of activity to be stored, but to let them carry out pattern completion at retrieval. Analytical estimates, derived for a simple model with discrete attractor states (Treves and Rolls 1992), suggest that this challenge can be met by afferent inputs with the characteristics and strength of the mossy fibers (Mori et al. 2007), but not by those conveyed by the perforant path to CA3, relayed by synapses that are presumably similar to recurrent synapses, but fewer in number. To make full contact with recordings of CA3 activity, however, the argument has to be generalized and applied to models in which CA3 units encode spatial representations, not just discrete attractor states.

CA3 as a memory store for spatial “charts”

Initially, the quantitative neural network analyses cited above had been formulated in terms of fully connected recurrent architectures and discrete memory states conceived—in the limit of no fluctuations—as points in a multidimensional space, in which each component corresponds to the firing rate or, in general, to the activity of one unit (Hopfield 1982). Thus, the salient spatial character of hippocampal memory correlates was provisionally neglected to take advantage of the formal models based on discrete attractor states. The very same autoassociator architecture, however, may subserve both the storage of discrete memories as point-like attractor states or of more complex memories, including continuous attractors, when network dynamics converges to fixed points that are not isolated, but continuously arranged on one or more low-dimensional manifolds embedded 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 1997) or of the head direction system (Skaggs et al. 1995). These models do not store information in long-term memory, and their
fixed points comprise a single (in these particular cases, one-dimensional) manifold. The multiple-chart model of Samsonovich and McNaughton (1997) demonstrated instead, in the context of a model for path integration, how one could conceive of fixed points organized in multiple two-dimensional continuous manifolds, each of which maps the animal’s position in a distinct environment.

**Context discrimination and exact localization from CA3 activity**

The model conceptually distinguishes two types of information that can both be extracted from CA3 activity: context discrimination (“which chart does the current activity pattern belong to?”) and exact localization within a context (“where in the chart?”). The distinction may be fuzzier for an animal who navigates among a succession of local contexts without clear boundaries, but it is very helpful at a theoretical level. For both types of information, part may be present, at any given time, in the inputs to CA3, and part may be retrieved there through pattern completion. Path integration and the predictive coding of future locations, to the extent that they are implemented within CA3, can be conceptualized as continuously sliding activity within the manifold spanned by a single chart, whereas the retrieval of the current context from partial cues corresponds to chart selection, i.e., to pattern completion in a slightly more general sense, in which the attractor is a full chart (Fig. 2A).

Within a chart, population activity on the surface of this attractor can therefore be conceived in the form of a “bump,” and it is then the position of the bump that codes for the position of the animal (Fig. 2B). On the surface of strictly continuous attractors there is no resistance to drift, hence, population activity is only marginally stable: A bump smoothly follows any variation of the input signal.

Individual cells may not be responsive anywhere on a chart, or they may have the characteristic place fields. Sharp (1991) demonstrated, using a simple neural network model with only a generic resemblance to hippocampal networks, how place cell-
like responses could arise from the association of different local views and how the direction-dependent place fields typical of one-dimensional environments may become direction insensitive in two dimensions through such associational mechanisms.

Place fields per se arise in networks of different architecture and plasticity, which are made to process spatial information (Zipser 1985; Treves et al. 1992; Burgess and O’Keefe 1996), so their mere appearance does not usefully constrain hypothetically hippocampal operations. Their relative quality, however, can give important insight into these operations, particularly as concerns exact localization within a context. One can assess the quality of place fields directly, e.g., visually from firing-rate maps, or for a quantitative appraisal of how encoding quality translates into network function, it is convenient to assess it in terms of localization measures.

As shown by Wilson and McNaughton (1993), one can easily deduce with reasonable accuracy the position of the rat in a recording box from a sufficient number of simultaneously recorded units. One constructs firing-rate vectors at any time $t$, as vectors $\mathbf{r}(t)$, where each component $r_i$ is the firing rate of one unit; then one may simply average all vectors expressed when the rat is in position $(x,y)$ to “wring” or “squeeze” out a continuous attractor. This also implies that such attractors are fractal, and, finally, by finding the best match for a vector at time $t$ among all templates, one can assign a decoded position to time $t$. More sophisticated procedures yield somewhat better performance. The very same decoding procedure can be applied to data obtained by simulating simple network models. Moreover, if a rat, real or simulated, has explored several environments, one may measure both the accuracy of decoding its position in each environment and of decoding which environment the rat is in, both with the same procedure and using multiple sets of templates, one per environment.

Fractured attractors tend to collapse

The mathematical analysis of attractor states applies, strictly speaking, only to networks with infinitely many cells whose place fields are arranged with infinite regularity to tile each chart. In this limit case there is no resistance to drift on the surface of a continuous attractor. This also implies that such attractors are susceptible to noise, which would push a bump away from its coding location. In a real system, however, beside noise there are also several sources of disorder: Even when there is only one map in memory, disorder arises from the finite size of the system, from the partial connectivity, from non-effective assignment of connection weights, etc. It was pointed out by Todyks and Sejnowski (1995) that the continuity of a one-dimensional “ring” attractor would be broken already by the moderate irregularity associated with assigning units to code random positions along the ring instead of equally spaced positions. In statistical physics terms, the free energy is not quite constant along the ring; it has valleys at slightly lower levels, where the units happen to be more concentrated, so that the network relaxes its activity in one of the discrete states at the bottom of the valleys rather than being indifferently moveable anywhere on the ring. To stabilize network activity at arbitrary positions requires, for example, sublinear activation functions (Stringer et al. 2002a) of the type that might be implemented by NMDA-receptor currents (Listman et al. 1998). The same phenomenon manifests itself in a different fashion if one attempts to create a continuous one-dimensional attractor not on a ring, which has no boundary, but on a linear segment, which has two extreme points. If equally spaced patterns are stored through a standard “Hebbian” learning rule, they need to be encoded with different weights, heavier at the extremes; otherwise, the extremes collapse onto the median point (Blumenfeld et al. 2006).

With two-dimensional would-be continuous attractors, the relevant case for the chart model, the effects of inhomogeneity in the distribution of place field centers and of boundary conditions are much stronger than in one dimension. Even with networks of considerable size, to establish reasonably smooth two-dimensional attractors, such that activity can settle at almost arbitrary positions, one needs special learning rules (Kali and Dayan 2000) or external factors such as position-dependent thresholds (Stringer et al. 2002b) or gain modulation (Roudi and Treves 2006). As a result, a bump, without further sensory input, would slowly drift away from its coding location and stabilize in a local or in the global minimum of the free-energy surface (Hamaguchi and Hatchett 2006). This drift may, however, be much slower than pattern completion. That is, efficient coding of a position is possible on the timescale between the attraction to the surface and before the bump drifts away by a reasonable amount (Fig. 2A). Interestingly, since the bump is sustained by the attraction to the surface, the bump does not disintegrate. When more then one chart is stored, disorder is increased. Attractor “collapse” is then accelerated slightly at each extra chart added. Close to storage capacity, suddenly all charts become unstable, that is, bumps drift away very fast, pattern completion may occur, and a few charts, or an attractor with no convergence is observed at all. Interestingly, such malfunctions may occur independently for different locations in an environment; thus, part of a chart may be recalled perfectly, while another part may be lost. These effects may be very difficult to assess in the real hippocampus, which normally operates under the influence of afferent inputs. Unless one designs ad hoc experiments to let place field activity sustain itself (Jarosiewicz and Skaggs 2004), it is much easier to understand the issue by studying retrieval dynamics through network simulations.

Simulations of a medium-size self-organizing hippocampal model demonstrate that, if ad hoc mechanisms are not introduced, rather than two-dimensional charts, one obtains “wrinkled” attractors, where activity can settle in only one or a few positions (Papp and Treves 2006). In these simulations, a virtual rat explores one or several virtual environments, realized as square boxes with toroidal boundary conditions (the two-dimensional equivalent of a ring), while a population of CA3 units develops representations of each environment (Treves 2004). If the unsupervised self-organization of a chart is produced solely by a “Hebbian” learning rule, place-cell-like responses emerge spontaneously in CA3, but many units end up having one of a few available fields, or no field: the corresponding attractor has collapsed. These are finite size effects, and they can be alleviated by scaling up the number of units and the connectivity in the simulated network (Fig. 2C), by constraining activity to be less sparse than experimentally observed (e.g., $a_{CA3} = 0.2$ rather than $a_{CA3} = 0.03$), and by keeping inhomogeneities to a minimum; but even with thousands of CA3 units, the resulting attractors remain far from the ideal charts realized in mathematical models (see also Fig. 3).

This “attractor collapse” scenario is particularly relevant for threshold-linear units with no saturation, which may model neuronal activity far from saturation. Simulations with effectively binary units, instead, with activity determined by a sigmoidal activation between 0 and 1, yield bumps of activity that are much more stable and show virtually no drift in the absence of inputs (G. Papp and A. Treves, unpubl.). In the sigmoidal model, however, most units in a stable state have activity close to 1 or 0 (that is, maximum or zero activity), and such an activity profile is incapable of describing real place-fields, where activity changes gradually over space (see also Roudi and Treves 2006).

How to iron out two-dimensional charts?

In the multiple chart model, exploration of a new environment must lead to the formation of a new chart. A number of questions...
then arise. How is the new chart laid out, ab initio or using some prewired connectivity? Are inputs from a population of grid cells useful? Can the process be aided by the dentate, as for discrete attractors? Is the final chart as smooth as the ideal concept of a continuous attractor implies? How many charts can possibly accumulate in a single recurrent network? These are very much issues of current research, and in the following we only discuss some of the results already obtained within the modeling approach.

In relation to the last question, discrete memory states models point at sparsity as the crucial representational parameter that influences memory capacity (beside the anatomical one of the connectivity per unit), as illustrated by Equation 1 above. Hippocampal space-related activity lends itself to the measurement of sparsity values, e.g., defined as the square of the mean firing rate across a population of units at a given instant, divided by the mean of the square firing rate (Treves and Rolls 1991). Applying such measure to individual cells recorded in a freely foraging task (courtesy of Jill and Stefan Leutgeb, Centre for the Biology of Memory, NTNU, Trondheim, Norway) one obtains values for CA3 in the sparse range, \( a_{CA3} \approx 0.02-0.06 \) (Papp and Treves 2007). The storage capacity of a multichart recurrent autoassociator was analyzed by Battaglia and Treves (1998), who extracted a simple rule-of-the-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, capacity-wise, to as many discrete attractor states as there are locations in the environment, for which the activity vectors are pairwise decorrelated. If the two-dimensional environment is represented by place-cell-like units, which are quiescent outside of 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 some dozen typical CA3 fields, say, “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 of the size of a typical recording box that can be held simultaneously in the network is theoretically limited by the critical value

\[
P_{charts} = 0.1 C/ln(1/a)
\]  

(see Figs. 1 and 2 in Battaglia and Treves 1998). This is still a huge number, of the order of several hundreds, given the recurrent connectivity of the CA3 network in rodents (Amaral et al. 1990). 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 points attractors to cover, as effectively smaller tiles, the whole environment. This chart capacity again respects the associative memory theoretical upper bound alluded to above, 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). It remains to be seen, however, whether in practice the upper bound can be approached.

The challenge of confronting grid cells

In the simulations of Figure 2, the inputs to CA3 units were from other units that themselves had place-field-like properties. One may expect that introducing a model of mEC grid cells as an input station to CA3 would help produce better continuous attractors, given the regularity of experimentally observed grid fields and the accurate localization they allow (Fyhn et al. 2004; Hafting et al. 2005). Taken as an isolated mEC-CA3 feedforward network, the transformation from grid cells to place cells is similar to a two-dimensional Fourier transform, requiring only a few grid cells of different spatial frequencies, but otherwise straightforward—if the grid fields are precisely aligned and the feedforward connection weights are assigned precalculated values (Solstad et al. 2006). In fact, retrieval dynamics proceed smoothly from mEC-like grid inputs if the CA3 representation has earlier been established with the aid of strong DG inputs as in Figure 3.

If the weights have to emerge from a self-organizing process,

\[a_{CA3} \approx 0.02-0.06\] (Papp and Treves 2007). The storage capacity of a multichart recurrent autoassociator was analyzed by Battaglia and Treves (1998), who extracted a simple rule-of-the-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, capacity-wise, to as many discrete attractor states as there are locations in the environment, for which the activity vectors are pairwise decorrelated. If the two-dimensional environment is represented by place-cell-like units, which are quiescent outside of 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 some dozen typical CA3 fields, say, “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 of the size of a typical recording box that can be held simultaneously in the network is theoretically limited by the critical value

\[
P_{charts} = 0.1 C/ln(1/a)
\]  

(see Figs. 1 and 2 in Battaglia and Treves 1998). This is still a huge number, of the order of several hundreds, given the recurrent connectivity of the CA3 network in rodents (Amaral et al. 1990). 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 points attractors to cover, as effectively smaller tiles, the whole environment. This chart capacity again respects the associative memory theoretical upper bound alluded to above, 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). It remains to be seen, however, whether in practice the upper bound can be approached.

The challenge of confronting grid cells

In the simulations of Figure 2, the inputs to CA3 units were from other units that themselves had place-field-like properties. One may expect that introducing a model of mEC grid cells as an input station to CA3 would help produce better continuous attractors, given the regularity of experimentally observed grid fields and the accurate localization they allow (Fyhn et al. 2004; Hafting et al. 2005). Taken as an isolated mEC-CA3 feedforward network, the transformation from grid cells to place cells is similar to a two-dimensional Fourier transform, requiring only a few grid cells of different spatial frequencies, but otherwise straightforward—if the grid fields are precisely aligned and the feedforward connection weights are assigned precalculated values (Solstad et al. 2006). In fact, retrieval dynamics proceed smoothly from mEC-like grid inputs if the CA3 representation has earlier been established with the aid of strong DG inputs as in Figure 3.

If the weights have to emerge from a self-organizing process,
even the most effective algorithm, a variant of Independent Component Analysis subject to a sparsity constraint, was found to produce rather implausible place fields even when summing from 100 grid units (Franzius et al. 2007). The main difficulty, for the algorithm that has to structure the weights appropriately, is how to suppress the periodicity inherent in the grid fields to lead to a single-peaked place field: A competitive learning algorithm may only reduce the mean number of peaks of the output units (Rolls et al. 2006). The difficulty is alleviated when the algorithm has available the input of an increasing number of grid units, of the order of thousands. Even in this situation, however, the self-organization of the feedforward weights is a delicate process, easily disrupted by the concurrent effect of any recurrent weights. In the presence of massively recurrent connectivity, as in CA3, no one has been able to demonstrate so far how direct grid cell inputs could be the driving force for the establishment of a new chart—all the more unlikely, if the recurrent connections are structured by the earlier storage of other charts, which interfere with the new one. It is reasonable to expect that combining grid cell inputs with nonperiodic spatial inputs representing, e.g., the activity of cells in lateral entorhinal cortex (lEC), should help with the formation of place fields; but evidence about the spatial correlates of IEC activity is too fragmentary yet to effectively constrain computational models.

The dentate gyrus as a chart preprocessor

The recent characterization of multipeaked place fields in the dentate gyrus (Leutgeb et al. 2007) provides a breakthrough for the development of models of chart formation in CA3. The observed fields resemble those produced by self-organization of feedforward inputs from grid-like units (Rolls et al. 2006; Franzius et al. 2007), redefining the feedforward models as relevant for studying granule cell activity and its changes after different manipulations. At the same time, the question arises as to whether one has been able to demonstrate so far how direct grid cell inputs could be the driving force for the establishment of a new chart—all the more unlikely, if the recurrent connections are structured by the earlier storage of other charts, which interfere with the new one. It is reasonable to expect that combining grid cell inputs with nonperiodic spatial inputs representing, e.g., the activity of cells in lateral entorhinal cortex (lEC), should help with the formation of place fields; but evidence about the spatial correlates of IEC activity is too fragmentary yet to effectively constrain computational models.

What computational models can attempt to resolve is whether such inputs can, given an appropriate synaptic structure with MF characteristics, overcome the unwelcome effects of the prevailing recurrent CA3 circuitry, including the “wrinkling” and eventual collapse of individual charts and the coalescence of new charts with previous ones.

With simulations, one can investigate the emergence of new charts in models with contrasting architecture, e.g., with and without a layer modeling the dentate gyrus. Figure 4 shows the results of simplified simulations, in which the same network was trained as a virtual rat explored a new environment. The DG layer, if present, is modeled here with single-field granule units and one-to-one detonator synapses to CA3 units, extrapolating somewhat current experimental evidence on the strength and indirect suppressive effect of MF activity (Mori et al. 2007), and is active only during training. At testing, DG is turned off, and CA3 cells, activated by the model perforant path and under the influence of recurrent collaterals, both modified during training, show scattered spatial responses before training, which self-organize into smoother fields after training (Fig. 4).

The DG teaching input modifies, in this simulation, both the perforant path and the recurrent collateral weights, and both contribute to the response properties of CA3 units. As the perforant path, which at testing relays only a partial cue, is made to gradually fade over 10 × 12.5-msec iterations, toward the end of the iteration cycle the cue is largely completed by the collateral effect (Fig. 3).

Note that in these preliminary simulations, MF synapses are one-to-one, and DG units have single-peaked fields, unlike those observed by Leutgeb et al. (2007). Still, DG units have spatially continuous firing rate maps, which is the crucial element to generate continuous representations of space, and which is beyond the scope of the original Treves and Rolls (1992) argument. We suggest that a thorough quantitative analysis of information storage in a model CA3 network, operating with and without dentate gyrus, is needed to assess again any information theoretic advan-

![Figure 4](image-url)  
**Figure 4.** Development of CA3 place fields with and without a model DG array. (A) Training the model in new environment with DG inputs (solid line) increases percent correct localization relative to training without DG (dashed line). Learning for one epoch (corresponding to ~10 min for a real rat) in both cases increases the accuracy afforded by decoding the entire population of 4900 CA3 units; further training without DG, however, is detrimental. (B) Three examples of firing rate maps in the model CA3 population are shown before and after training. Without DG, CA3 fields do not develop single peaks (note that single peaks occurring across the toroidal boundaries appear as double when the torus is displayed as a square box). Note also the refinement in the fields with the second epoch of training.
tage in forming new representations, this time in the form of charts of place-cell-like units. Unfortunately, since the very two-dimensional nature of charts makes a simplified mathematical analysis of information storage like the one in Treves and Rolls (1992) difficult to work out, computer simulations at present offer a practical approach to this issue.

In our simulations, considering one time-step to be 12.5 msec, attractors tend to collapse over several seconds (Fig. 2). Some deterioration in the accuracy of localization may already appear within the equivalent of 100-150 msec (Fig. 3). In rodents, theta-oscillations pace activity over a similar time scale, and it might just be that the collapse phenomenon, because of the typically sustained afferent inputs (e.g., from mEC units) and because of the limited time for convergence, is not so relevant to the real hippocampus. Theta oscillations are, however, characteristic of exploratory behavior, usually associated with memory encoding in the CA3 network, not retrieval; and the correspondence between real hippocampal dynamics in different behavioral conditions and the simplified dynamics of simulated models must be elaborated much better, probably with more complicated models, before dismissing the implications of attractor collapse.

At any rate, the collapse of attractors is intimately related to the recurrent nature of the CA3 network. If one makes partial "lesions" to the recurrent connections, attractors become smoother. At the same time, however, one loses the benefits of the collateral effect (Fig. 3). A possible solution to this problem, seemingly "invented" by mammals as mentioned above, might be the insertion of a population of units right after CA3 as a post-processor, i.e., the CA1 feedforward network. In simulations to be reported elsewhere (Papp and Treves 2007) we observe that CA1 processing may smooth the representation retrieved from CA3, thereby increasing its spatial information content (Treves 2004).

Morphing memories and morphing environments
Do discrete boundaries between attractor states survive manipulations in which external correlates are intentionally interpolated, i.e., morphed, between pre-established representations? In simulations of the morphing experiments of Wills et al. (2005) and Leutgeb et al. (2005b), we aimed first to test the idea that correlations between encoded patterns may lead to a linearization of responses along the morphing sequence, as observed by Leutgeb et al. (2005b) in contrast to Wills et al. (2005) (Papp and Treves 2005). Second, in the model, we aimed to assess the relevance of the ratio of recurrent to afferent connections in driving recall dynamics. The first issue is relevant, as a crucial difference between the two experiments involved the degree of orthogonalization of the spatial representations originally established (whether they led to complete remapping or not, see Leutgeb et al. 2005a). The second issue is relevant, in relation to the hypothesized effect of ACh, even if this is not explicitly modeled in the simulations. We proceed by first using simple, discrete patterns and then two-dimensional charts.

Morphing discrete patterns
The model simulated in this and the following section includes two layers: "entorhinal cortex," serving as the input, and CA3, the actual attractor network. Each of them is comprised of 3600 threshold-linear units (Treves and Rolls 1991), arranged on a 60 x 60 grid with periodic boundary conditions. Connectivity is partial: 240 units randomly chosen in the EC array connect to any CA3 unit, which also receives from a random assortment of 400 fellow CA3 units. Discrete patterns or spatial charts are not the product of extensive training, but rather they are "assigned" to the CA3 network in a crude model of one-shot associative learning as in the Hopfield (1982) model. A simple Hebbian learning rule is used, as in Treves (2004). Patterns are set also in the EC layer, and associated with those in CA3 with the same one-shot learning rule. During testing, a full input pattern is presented in EC, and the network is updated for several hundred timesteps.

Different input patterns in EC, e.g., A and B (representing, for example, the circle and square box used in the experiments), were created by scrambling EC units among themselves. Morphs were then set up by setting the activity of a fraction of EC units to their activity in one pattern and the remaining cells to their activity in another pattern. Changes from one full pattern to the first morph, between the morphs, and between the last morph and the second full pattern was set to be equispaced. Correlated initial patterns were created with an incomplete scrambling; thus, a number of EC units had, in this case, the same representation for A and B, whereas the others were allowed to have different representation.

The storage capacity of the network was first established and was found to be around 60 independent patterns. To stay well below storage capacity, only four discrete patterns were used in the morphing simulation. Results were mapped as a function of two parameters, the recurrent-to-afferent strength and the degree of correlation between A and B. In Figure 5A four representative examples are shown for these two parameters. As in Leutgeb et al. (2005b), five intermediate morphs were used, denoted as AAAB, AAB, AB, ABB, and ABBB, respectively. The reference case, with "normal" recurrent strength and no correlation between A and B, is shown in panel 3 and corresponds to a sudden transition between the two discrete attractors for A and for B. Decreasing the recurrent strength by a factor of 10 (Fig. 5A, panel 1), the transition becomes linear. A very slight trace of pattern completion can be still detected (as overlap scores for AAAB and AAB above the line connecting those for A and B), showing that although the strength of recurrent weights is decreased, they are still not completely ineffective. Introducing a correlation of 0.3 (that is, 30% of EC units are kept to have the same representation in A and B) in the case with strong recurrent weights (panel 4) strongly increases the overlap between the representation of A and B, and the sudden transition is lost, replaced by an accelerating decrease in overlap scores over the morphing sequence. Thus, introducing correlations between attractors has a similar effect to reducing recurrent ratios. Finally, again decreasing the effect of recurrent connections, while also keeping the elevated correlation (Fig. 5A, panel 2), a smooth linear transformation appears between similar attractor states, with overlap scores that remain high even between A and B.

Further simulations show that an even higher degree of correlation leads to a complete collapse of the memory states for A and B. The collapse, much like the collapse within a chart (described in the previous section) results from the action of recurrent collateral connections, as shown by the finding that when removing such connections altogether, the representations of A and B remain different and the transition over the morphing sequence linear. Like the collapse within a chart, it will not manifest itself when the network is sufficiently driven, that is, steered by afferent inputs. Note that recurrent weights in these simulations reflect the storage of only the extreme patterns, unlike the model analyzed by Blumenfeld et al. (2006).

Morphing charts
Simulations with two-dimensional representations were conducted with the same procedure as for the discrete attractor case, however in a more complex situation, where spatial charts are
stored in the network instead of discrete patterns. To attain the larger capacity needed to store charts, a network comprised of 4900 CA3 units arranged on a 70 x 70 grid was used, with 400 afferent inputs and 1200 recurrent connections each. The network was found to store up to eight different charts before a clear capacity breakdown, and four charts were used in the morphing simulations to stay below storage capacity. Using periodic boundary conditions, each chart was arranged on a torus, and a scrambling procedure was used on the torus to obtain morphed environments. Connection weights were precalculated as a sum of contributions from the different charts stored, in each decreasing as a function of the distance between the position of the presynaptic and postsynaptic unit in the chart. Parameters were chosen through an extensive search to find conditions that would allow the slowest drift of population activity, i.e., the best stability in the chart.

In order to set up the chart, the centers of place fields in EC and in CA3 were assigned prior to setting up the weights. Each unit, both in EC and in CA3, was simply set to have a field in all of the environments (decreasing the number of units coding for an environment decreases the stability of the resulting memory, as measured by percent correct localization). Note, therefore, that for simplicity in these simulations, each EC unit was set to have one peak, unlike the simulations with grid units described earlier. The morphing effect discussed here is an effect endogenous to the recurrent CA3 network, and thus likely to emerge irrespective of the input pattern in EC.

In a case of recall with strong recurrents and no correlation (Fig. 5B, panel 3), the mean overlap for morphs AAAB and AAB show pattern completion toward morph A, whereas for ABB and ABBB the CA3 units tend to converge toward the representation of B. In middle morph, some locations are coded as more similar to those in A, while others to those in B. In morphs AAB and ABB, in fact, there are already some locations acquiring the representation of the extreme shape that is more distant in terms of the morphing sequence. Still, the representation in the majority of spatial locations switches abruptly between the morphs AAB and AB. Individual place fields mostly follow the same dynamics, with most often global remapping between morphs AAB and AB. Decreasing the strength of recurrent connections leads again to a linearization of the transition (panel 1). There is now little sign of pattern completion, however, unlike that seen with discrete attractors. Place fields change smoothly over the whole morphing sequence (data not shown). Again, that implies that a network
predominantly driven by external inputs, e.g., from mEC, is unlikely to show the features characterizing attractor dynamics, be they advantageous, like pattern completion, or disadvantageous, like attractor collapse.

Introducing correlation (panel 4) between the neural representations of A and B formed by CA3 also abolishes the sudden transition. As shown by the scatter of individual dots in panel 4, individual locations change their population code quasi independently over the morphing sequence, similar to experimental results by Leutgeb et al. (2005b). In the model, however, no unit acquired a place field only in some intermediate morphs, because in the model only the extremes A and B were used in setting connection weights, and in A and B each unit had a place field. Decreasing the relative strength of the recurrent connections further linearizes the morphing curve (panel 2), with the overlap between A and B remaining quite elevated, again consistent with the findings of Leutgeb et al. (2005b). Place field again changed smoothly.

A conclusion from these morphing simulations is that correlations in the environment, as well as between discrete stimuli represented in a strongly recurrent network, lead to attractor collapse, a different manifestation of the same phenomenon apparent with the storage of continuous attractors, as discussed above. The drift to discrete attractor locations within a would-be continuous attracting manifold is augmented when multiple manifolds are stored, by their collapsing onto each other. Both phenomena remain latent when the network is, broadly speaking, steered by afferent inputs: a weakly recurrent network is less subject to attractor collapse, but then population activity tends to track ongoing inputs anyway, resulting in essentially linear transitions along a morphing sequence. The self-organized formation of a single continuous attractor and the retrieval of pre-established representations from morphed inputs, therefore, both face the same challenge in the recurrent CA3 network of separating patterns that are separated in the inputs. When multiple spatial charts are to be self-organized from scratch, the difficulty for the CA3 network is redoubled, and further computational work may help assess to what extent the dentate gyrus may help meet the challenge.

Conclusions

As originally proposed by Marr (1971) and then by McNaughton and Morris (1987) and by Rolls (1989), the general theory of autoassociative memory networks provides a most useful model with which to gain insight into the information-processing operations performed by the hippocampal CA3 network. A comparison with the organization of the medial telencephalon in reptiles and birds, in fact, indicates in CA3, in contrast to other hippocampal subfields in mammals, the most preserved traits of the archetypical organization of primitive cortex. This suggests that in functional terms, too, the operation of CA3 may be the straightforward evolutionary derivative of an archetypical storage site for complex memories involving relations between disparate elements, and capable of autoassociative retrieval from arbitrary partial cues. The mathematical analysis introduced for the Hopfield (1982) model by Amit et al. (1987), as well as computer simulations on more realistic versions of the model, allow for a quantitative appraisal of how efficiently CA3 may function as an autoassociative memory network, in particular in terms of the crucial constraint of storage capacity. These approaches reveal that acquiring new memory representations in general, and acquiring spatial representations in particular, even more when derived from grid units, pose hard challenges to CA3. In mammals, meeting these challenges may have required the evolution of additional specialized circuitry, including at least the dentate gyrus as a preprocessing stage to CA3. It remains to be explored whether other, distinct evolutionary adaptations may have met similar challenges in the avian brain, given the intriguing suggestion that birds, unlike reptiles, appear to have evolved similar slow-wave sleep dynamics as mammals (Rattenborg 2007).

Acknowledgments

We thank all colleagues at the Centre for the Biology of Memory and Floris Wouterlood and Piet Hoogland at the VUMC, and two reviewers for their helpful suggestions.

References

Amaral, D.G. and Witter, M.P. 1989. The three-dimensional organization of the hippocampal formation: A review of anatomical data. Neuroscience 31: 571–591.

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

Amit, D.J. 1989. Modeling brain function. Cambridge UP, Cambridge, UK.

Amit, D.J., Gutfreund, H., and Sompolinsky, H. 1987. Statistical mechanics of neural networks near saturation. Ann. Phys. 173: 30–67.

Andersen, P. and Lyonning, Y. 1962. Interaction of various afferents on CA1 neurons and dentate granule cells. Colloq. Int. CNRS 107: 23–45.

Ariens-Kapper, C.U., Huber, G.C., and Crosby, E.C. 1936. The comparative anatomy of the nervous system of vertebrates, including man. Halter Publishing, New York.

Atoji, Y. and Wild, J.M. 2004. Fiber connections of the hippocampal formation and septum and subdivisions of the hippocampal formation in the pigeon as revealed by tract tracing and kainic acid lesions. J. Comp. Neurol. 475: 426–461.

Atoji, Y. and Wild, J.M. 2006. Anatomy of the avian hippocampal formation. Rev. Neurosci. 17: 3–15.

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

Bingman, V.P. and Sharp, P.E. 2006. Neuronal implementation of hippocampal-mediated spatial behaviour: A comparative evolutionary perspective. Behav. Cogn. Neurosci. Rev. 5: 80–91.

Bliss, T.V. and Lomo, 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.

Blenkenfeld, B., Preminger, S., Nagi, D., and Toodyks, M. 2006. Dynamics of memory representations in networks with novelty-facilitated synaptic plasticity. Neuron 52: 383–394.

Bratkenberg, V. and Schütz, A. 1991. Anatomy of the cortex. Springer Verlag, Berlin, Germany.

Burgess, N. and O’Keefe, J. 1996. Neuronal computations underlying the firing of place cells and their role in navigation. Hippocampus 6: 749–762.

Clayton, N.S., Bussey, T.J., and Dickinson, A. 2003. Can animals recall the past and plan for the future? Nat. Rev. Neurosci. 4: 685–691.

Colombo, M. and Broadbent, N. 2000. Is the avian hippocampus a functional homologue of the mammalian hippocampus? Neurosci. Biobehav. Rev. 24: 465–484.

Dubeldam, J.L. 1998. The neural substrate for ‘learned’ and ‘nonlearned’ activities in birds: A discussion of the organization of bulbil reticular premotor systems with side-lights on the mammalian situation. Acta Anat. (Basel) 163: 157–172.

Erichsen, J.T., Bingman, V.P., and Krebs, J.R. 1991. The distribution of neuropeptides in the dorsomedial telencephalon of the pigeon (Columbia livia): A basis for regional subdivisions. J. Comp. Neurol. 314: 478–492.

Fabre, H., Braun, K., Zschiedrath, W., and Scheich, H. 1989. System-specific distribution of zinc in the chick brain. A light- and electron-microscopic study using the Timm method. Cell Tissue Res. 258: 247–257.

Franzén, M., Vollgraf, R., and Wiskott, L. 2007. From grids to places. J. Comput. Neurosci. 22: 297–299.

Fyhn, M., Molden, S., Witter, M.P., Moser, E.I., and Moser, M.-B. 2004. Spatial representation in the entorhinal cortex. Science 305: 1258–1264.

Gardner-Medwin, A.R. 1976. The recall of events through the learning process. Behav. Cogn. Neurosci. Rev. 80: 383–394.

Gardner-Medwin, A.R. 1976. The recall of events through the learning process. Behav. Cogn. Neurosci. Rev. 80: 383–394.

Gardner-Medwin, A.R. 1976. The recall of events through the learning process. Behav. Cogn. Neurosci. Rev. 80: 383–394.

www.learnmem.org
Stephan, H. 1975. Allocortex. In *Handbuch der mikr. Anatomie des Menschen*, (ed. W. Bargmann) Vol. 4, Teil 9. pp. 998. Springer, Berlin, Heidelberg, New York.

Streit, G.F. 2005. *Principles of Brain Evolution*. Sinauer Associates, Sunderland, MA.

Stringer, S.M., Trappenberg, T.P., Rolls, E.T., and de Araujo, I.E.T. 2002a. Self-organizing continuous attractor networks and path integration: One-dimensional models of head direction cells. *Network* 13: 217–242.

Stringer, S.M., Rolls, E.T., Trappenberg, T.P., and de Araujo, I.E.T. 2002b. Self-organizing continuous attractor networks and path integration: Two-dimensional models of place cells. *Network* 13: 429–446.

Székely, A.D. and Krebs, J.R. 1996. Efferent connectivity of the hippocampal formation of the zebra finch (*Taeniopygia guttata*): An anterograde pathway tracing study using *Phaseolus vulgaris* leucoagglutinin. *J. Comp. Neurol.* 368: 198–214.

ten Donkelaar, H.J. 2000. Major events in the development of the forebrain. *Eur. J. Morphol.* 38: 301–308.

Treves, A. 2004. Computational constraints between retrieving the past and predicting the future, and the CA3-CA1 differentiation. *Hippocampus* 14: 539–556.

Treves, A. and Rolls, E.T. 1991. What determines the capacity of autoassociative memories in the brain? *Network* 2: 371–397.

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

Tsodyks, M. and Sejnowski, T. 1995. Associative memory and hippocampal place cells. *Int. J. Neural Syst.* 6: 81–86.

Ulinsky, P.S. 1990a. The cerebral cortex of reptiles. In *Cerebral Cortex: Comparative Structure and Evolution of Cerebral Cortex* (eds. E.G. Jones and A. Peters), Vol. 8A, pp. 139–215. Plenum Press, New York.

Ulinsky, P.S. 1990b. Nodal events in the forebrain evolution. *Neth. J. Zool.* 40: 213–240.

Unal, B., Bradley, P.M., Sahin, B., Canan, S., Aslan, H., and Kaplan, S. 2002. Estimation of numerical density and mean synaptic height in chick hippocampus 24 and 48 hours after passive avoidance training. *Brain Res. Dev. Brain Res.* 136: 135–144.

Wierszko, A. and Ball, G.F. 1993. Long-term potentiation in the avian hippocampus does not require activation of the N-methyl-D-aspartate (NMDA) receptor. *Synapse* 13: 173–178.

Wills, T.J., Lever, C., Cacucci, F., Burgess, N., and O’Keefe, J. 2005. Attractor dynamics in the hippocampal representation of the local environment. *Science* 308: 873–876.

Wyss, J.M. and Sripanidkulchai, K. 1983. The indusium griseum and anterior hippocampal continuation in the rat. *J. Comp. Neurol.* 219: 251–271.

Zipser, D. 1985. A computational model of hippocampal place fields. *Behav. Neurosci.* 99: 1006–1018.

Received July 1, 2007; accepted in revised form September 5, 2007.
The CA3 network as a memory store for spatial representations

Gergely Papp, Menno P Witter and Alessandro Treves

Learn. Mem. 2007, 14:
Access the most recent version at doi:10.1101/lm.687407

References
This article cites 85 articles, 16 of which can be accessed free at:
http://learnmem.cshlp.org/content/14/11/732.full.html#ref-list-1

License

Email Alerting Service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.