J. van Pelt, M.A. Corner H.B.M. Uylings and F.H. Lopes da Silva (Eds.)
Progress in Brain Research, Vol 102

CHAPTER 22

Neural networks in the brain involved in memory and recall

Edmund T. Rolls and Alessandro Treves*

Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford, OX1 3UD, U.K.

Introduction

Damage to the hippocampus and related structures leads to anterograde amnesia, i.e. an inability to form many types of memory. Old memories are relatively spared. Recent memories, formed within the last few weeks or months, may be impaired (Squire, 1992). The learning tasks that are impaired include spatial and some non-spatial tasks in which information about particular episodes, such as where a particular object was seen, must be remembered. Further, some hippocampal neurons in the monkey respond to combinations of visual stimuli and places where they are seen (Rolls, 1990a,b, 1991; Rolls and O'Mara, 1993). It is also known that inputs converge into the hippocampus, via the adjacent parahippocampal gyrus and entorhinal cortex, from virtually all association areas in the neocortex, including areas in the parietal cortex concerned with spatial function, temporal areas concerned with vision and hearing, and the frontal lobes (Fig. 1). An extensively divergent system of output projections enables the hippocampus to feed back into most of the cortical areas from which it receives inputs. On the basis of these and related findings the hypothesis is suggested that the importance of the hippocampus in spatial and other memories is that it can rapidly form 'episodic' representations of information originating from many areas of the cerebral cortex, and act as an intermediate term buffer store. In this paper, analyses of how the architecture of the hippocampus could be used as such a buffer store are considered, and then a hypothesis on how recent memories could be recalled from this store back into the cerebral cortical association areas, to be used as needed in the formation of long-term memories, is presented (see Marr, 1971; Rolls, 1989a,b, 1991; Treves and Rolls, 1994).

The hippocampus

Hippocampal CA3 circuitry (see Fig. 1)

Projections from the entorhinal cortex reach the granule cells (of which there are 10^6 in the rat) in the dentate gyrus (DG) via the perforant path (pp). The granule cells project to CA3 cells via the mossy fibres (MF), which provide a *sparse* but possibly powerful connection to the 3×10^5 CA3 pyramidal cells in the rat. Each CA3 cell receives approximately 50 mossy fibre inputs, so that the sparseness of this connectivity is thus 0.005%. By contrast, there are many more—pos-

* Present address: S.I.S.S.A. - Biophysics, via Beirut 2–4, 34013 Trieste, Italy.

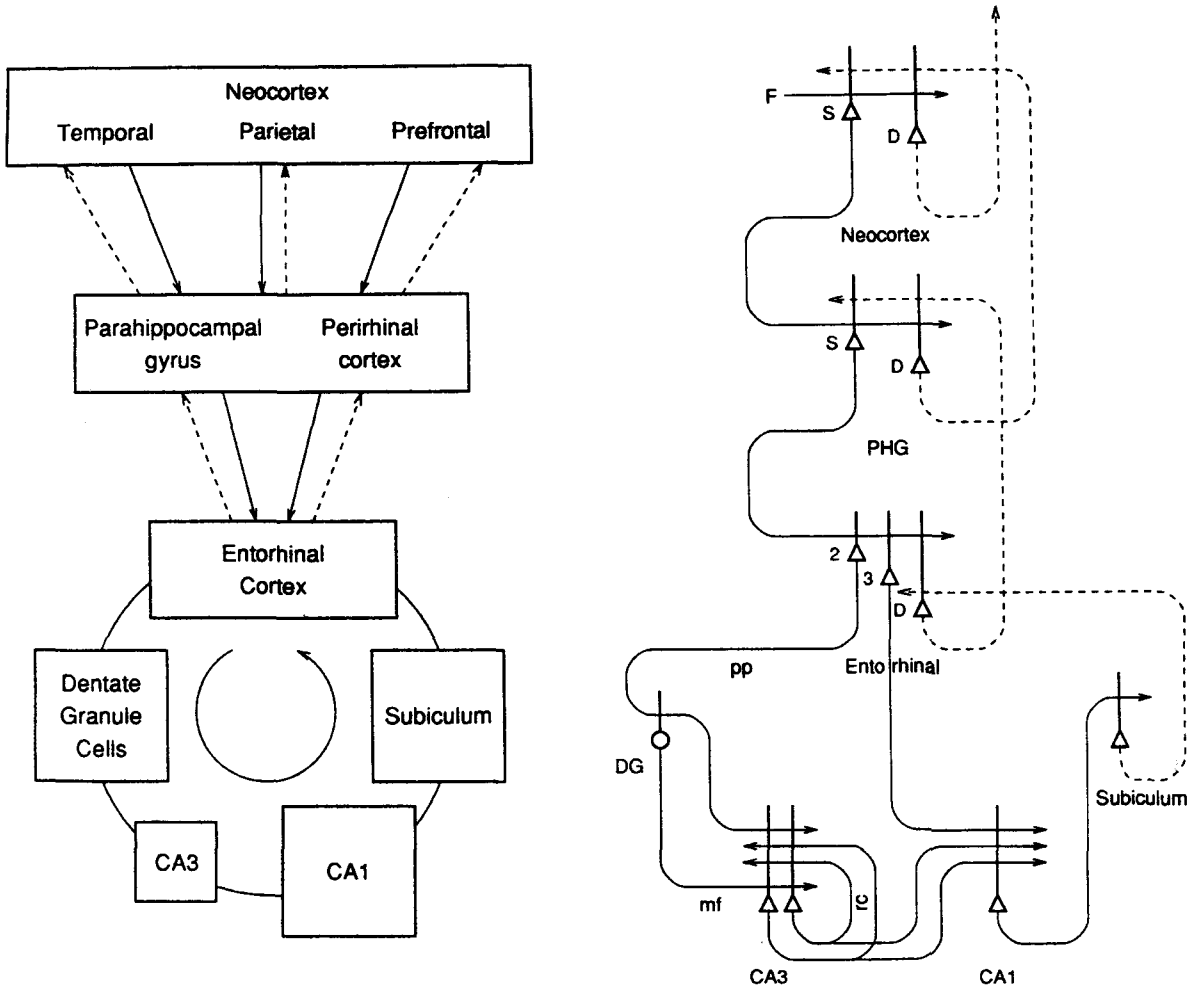

Fig. 1. Forward connections (solid lines) from areas of cerebral association neocortex via the parahippocampal gyrus and perirhinal cortex, and entorhinal cortex, to the hippocampus, and back-projections (dashed lines) via the hippocampal CA1 pyramidal cells, subiculum, and parahippocampal gyrus to the neocortex. There is great convergence in the forward connections down to the single network implemented in the CA3 pyramidal cells, and great divergence again in the back-projections. Left: block diagram. Right: more detailed representation of some of the principal excitatory neurons in the pathways. Abbreviations—D: deep pyramidal cells; DG: dentate granule cells; F: forward inputs to areas of the association cortex from preceding cortical areas in the hierarchy; mf: mossy fibres; PHG: parahippocampal gyrus and perirhinal cortex; pp: perforant path; rc: recurrent collateral of the CA3 hippocampal pyramidal cells; S: superficial pyramidal cells; 2: pyramidal cells in layer 2 of the entorhinal cortex; 3: pyramidal cells in layer 3 of the entorhinal cortex. The thick lines above the cell bodies represent the dendrites.

sibly weaker—direct perforant path inputs onto each CA3 cell, in the rat of the order of 40×10^3. The largest number of synapses (about 1.2×10^4 in the rat) on the dendrites of CA3 pyramidal cells is, however, provided by the (recurrent) axon collaterals of CA3 cells themselves (rc). It is remarkable that the recurrent collaterals are distributed to other CA3 cells throughout the hippocampus (Amaral and Witter, 1989; Amaral et al., 1990; Ishizuka et al., 1990), so that effectively

the CA3 system provides a single network, with a connectivity of approximately 4% between the different CA3 neurons.

CA3 as an autoassociation memory

Many of the synapses in the hippocampus show associative modification as shown by long-term potentiation, and this synaptic modification appears to be involved in learning (Morris, 1989). On the basis of the evidence summarized above, Rolls (1987, 1989a,b,c, 1990a,b, 1991) has suggested that the CA3 stage acts as an autoassociation memory which enables episodic memories to be formed and stored for an intermediate term in the CA3 network, and that subsequently the extensive recurrent collateral connectivity allows for the retrieval of a whole representation to be initiated by the activation of some small part of the same representation (the cue). We have therefore performed quantitative analyses of the storage and retrieval processes in the CA3 network (Treves and Rolls, 1991, 1992). We have extended previous formal models of autoassociative memory (see Amit, 1989) by analysing a network with graded response units, so as to represent more realistically the continuously variable rates at which neurons fire, and with incomplete connectivity (Treves, 1990; Treves and Rolls, 1991). We have found that in general the maximum number p_{max} of firing patterns that can be (individually) retrieved is proportional to the number C^{RC} of (associatively) modifiable RC synapses per cell, by a factor that increases roughly with the inverse of the sparseness a of the neuronal representation. The sparseness is defined as

$$a = \frac{< \eta >^2}{< \eta^2 >} \tag{1}$$

where $< \cdot >$ denotes an average over the statistical distribution characterizing the firing rate η of each cell in the stored patterns. Approximately,

$$P_{max} \simeq \frac{C^{RC}}{a \ln(1/a)} k \tag{2}$$

where k is a factor that depends weakly on the detailed structure of the rate distribution, on the connectivity pattern, etc., but is roughly in the order of 0.2–0.3 (Treves and Rolls, 1991).

The main factors that determine the maximum number of memories that can be stored in an autassociative network are thus the number of connections on each neuron devoted to the recurrent collaterals, and the sparseness of the representation. For example, for $C^{RC} = 12{,}000$ and $a = 0.02$ (realistic estimates for the rat), p_{max} is calculated to be approximately 36,000.

We have also indicated how to estimate I, the total amount of information (in bits per synapse) that can be retrieved from the network. I is defined with respect to the information i_p (in bits per cell) contained in each stored firing pattern, by subtracting the amount i_l lost in retrieval and multiplying by p/C^{RC}:

$$I \equiv \frac{p}{C^{RC}}(i_p - i_l). \tag{3}$$

The maximal value I_{max} of this quantity was found (Treves and Rolls, 1991) to be in several interesting cases around 0.2–0.3 bits per synapse, with only a mild dependency on parameters such as the sparseness of coding a.

The requirement of the input systems to CA3 for the efficient storage of new information

By calculating the amount of information that would end up being carried by a CA3 firing pattern produced solely by the perforant path input and by the effect of the recurrent connections, we have been able to show (Treves and Rolls, 1992) that an input of the perforant path type, alone, is unable to direct efficient information storage. Such an input is too weak, it turns out, to drive the firing of the cells, as the 'dynamics' of the network is dominated by the randomizing effect of the recurrent collaterals. This is the manifestation, in the CA3 network, of a general problem affecting storage (i.e. learning) in *all* autoassociative memories. The problem arises when the system is considered to be activated by a set of input axons making synaptic connections that have to

compete with the recurrent connections, rather than having the firing rates of the neurons artificially clamped into a prescribed pattern.

In an argument developed elsewhere, we hypothesize that the mossy fibre inputs force efficient information storage by virtue of their strong and sparse influence on the CA3 cell firing rates (Rolls, 1989a,b; Treves and Rolls, 1992).

A different input system is needed to trigger retrieval

An autoassociative memory network needs afferent inputs also in the other mode of operation, i.e. when it retrieves a previously stored pattern of activity. We have shown (Treves and Rolls, 1992) that if the cue available to initiate retrieval is rather small, one needs a large number of associatively modifiable synapses. The number needed is of the same order as the number of concurrently stored patterns p. For such reasons we suggest that the direct perforant path system to CA3 (see Fig. 1) is the one involved in relaying the cues that initiate retrieval.

The dentate granule cells

The theory is developed elsewhere that the dentate granule cell stage of hippocampal processing which precedes the CA3 stage acts to produce during learning the sparse yet efficient (i.e. non-redundant) representation in CA3 neurons which is required for the autoassociation to perform well (Rolls, 1989a,b,c, 1991; see also Treves and Rolls, 1992). One way in which it may do this is by acting as a competitive network to remove redundancy from the inputs producing a more orthogonal, sparse, and categorized set of outputs (Rolls, 1987, 1989a,b,c, 1990a,b). A second way arises because of the very low contact probability in the mossy fibre–CA3 connections, which helps to produce a sparse representation in CA3 (Treves and Rolls, 1992). A third way is that the powerful dentate granule cell–mossy fibre input to the CA3 cells may force a new pattern of firing onto the CA3 cells during learning.

The CA1 organization

The CA3 cells are connected to the CA1 cells. It is suggested that the CA1 cells, given the separate parts of each episodic memory which must be separately represented in CA3 ensembles, can allocate neurons, by competitive learning, to represent at least larger parts of each episodic memory (Rolls, 1987, 1989a,b,c, 1990a,b). This implies a more efficient representation, in the sense that when eventually after many further stages, neocortical neuronal activity is recalled (as discussed below), each neocortical cell need not be accessed by all the axons carrying each component of the episodic memory as represented in CA3, but instead by fewer axons carrying larger fragments.

Dynamics and the temporal dimension

The analysis described above of the capacity of a recurrent network such as the CA3 considered steady state conditions of the firing rates of the neurons. The question arises of how quickly the recurrent network would settle into its final state. With reference to the CA3 network, how long does it take before a pattern of activity, originally evoked in CA3 by afferent inputs, becomes influenced by the activation of recurrent collaterals? In a more general context, recurrent collaterals between the pyramidal cells are an important feature of the connectivity of the cerebral neocortex. How long would it take these collaterals to contribute fully to the activity of cortical cells. If these settling processes took in the order of hundreds of milliseconds, they would be much too slow to contribute usefully to cortical activity, whether in the hippocampus or the neocortex (Rolls, 1992). As this question is crucial to understanding cortical function, we have analysed the time taken for neocortical pyramidal cells in the inferior temporal cortex of macaques to settle into a stable pattern of firing in response to a visual stimulus. We have found that within 20–40 msec of first firing, that is within 1–3 spikes from the fastest firing cells, single neurons have settled into stable firing rates which provide approximately 50% of the maximal information which can be obtained from the spike train of the neuron (Tovee et al., 1993). Could recurrent collaterals contribute to settling as rapid as this?

A partial answer to this question can be in-

ferred from a recent theoretical development based on the analysis of the collective dynamical properties of realistically modelled neuronal units (Treves, 1993; Treves et al., 1994). The method incorporates the biophysical properties of real cell membranes, and considers the dynamics of a network of integrate-and-fire neurons, laterally connected through realistically modelled synapses. The analysis indicates that the model network will attain a stable distribution of firing rates over time scales determined essentially by synaptic and intrinsic conductance inactivation times. Some of these (e.g. the conductance time constants associated with excitatory synapses between pyramidal cells) are very short, less than 10 msec, implying that the activation of recurrent collaterals between pyramidal cells will contribute to determine the overall firing pattern within a period of a very few tens of msec (for further details see Treves, 1993; Treves et al., 1994). With respect to the CA3 network, the indication is thus that retrieval would be rapid, indeed fast enough for it to be biologically plausible.

Backprojections to the neocortex

The hippocampus as a buffer store

It is suggested that the hippocampus is able to recall the whole of a previously stored episode for a period of days, weeks or months after the episode, when even a fragment of the episode is available to start the recall. This recall from a fragment of the original episode would take place particularly as a result of completion produced by the autoassociation implemented in the CA3 network. It would then be the role of the hippocampus to reinstate in the cerebral neocortex the whole of the episodic memory. The cerebral cortex would then, with the whole of the information in the episode now producing firing in the correct sets of neocortical neurons, be in a position to incorporate the information in the episode into its long-term store in the neocortex.

We suggest that during recall, the connections from CA3 via CA1 and the subiculum would allow activation of at least the pyramidal cells in the deep layers of the entorhinal cortex (see Fig. 1). These neurons would then, by virtue of their backprojections to the parts of the cerebral cortex that originally provided the inputs to the hippocampus, terminate in the superficial layers of those neocortical areas, where synapses would be made onto the distal parts of the dendrites of the cortical pyramidal cells (see Rolls, 1989a,b,c).

Our understanding of the architecture with which this would be achieved is shown in Fig. 1. The feed-forward connections from association areas of the cerebral neocortex (solid lines in Fig. 1), show major convergence as information is passed to CA3, with the CA3 autoassociation network having the smallest number of neurons at any stage of the processing. The back-projections allow for divergence back to neocortical areas. The way in which we suggest that the back-projection synapses are set up to have the appropriate strengths for recall is as follows (see also Rolls, 1989a,b). During the setting up of a new episodic memory, there would be strong feed-forward activity progressing towards the hippocampus. During the episode, the CA3 synapses would be modified, and via the CA1 neurons and the subiculum, a pattern of activity would be produced on the back-projecting synapses to the entorhinal cortex. Here the back-projecting synapses from active back-projection axons onto pyramidal cells being activated by the forward inputs to the entorhinal cortex would be associatively modified. A similar process would be implemented at preceding stages of the neocortex, that is in the parahippocampal gyrus/perirhinal cortex stage, and in association cortical areas.

Quantitative constraints on the connectivity of back-projections

How many back-projecting fibres does one need to synapse on any given neocortical pyramidal cell, in order to implement the mechanism outlined above? Clearly, if the theory were to produce a definite constraint of the sort, quantitative anatomical data could be used for verification or falsification.

Consider a polysynaptic sequence of back-projecting stages, from hippocampus to neocortex, as a string of simple (hetero-)associative memories

in which, at each stage, the input lines are those coming from the previous stage (closer to the hippocampus). Implicit in this framework is the assumption that the synapses at each stage are modifiable and have been indeed modified at the time of first experiencing each episode, according to some Hebbian associative plasticity rule. A plausible requirement for a successful hippocampo-directed recall operation, is that the signal generated from the hippocampally retrieved pattern of activity, and carried backwards towards the neocortex, remain undegraded when compared with the noise due, at each stage, to the interference effects caused by the concurrent storage of other patterns of activity on the same back-projecting synaptic systems. That requirement is equivalent to that used in deriving the storage capacity of such a series of heteroassociative memories, and it was shown in Treves and Rolls (1991) that the maximum number of independently generated activity patterns that can be retrieved is given, essentially, by the same formula as (2) above

$$p \simeq \frac{C}{a\ln(1/a)}k' \qquad (2')$$

where, however, a is now the sparseness of the representation at any given stage, and C is the average number of (back-)projections each cell of that stage receives from cells of the previous one. (k' is a similar slowly varying factor to that introduced above.) If p is equal to the number of memories held in the hippocampal buffer, it is limited by the retrieval capacity of the CA3 network, p_{max}. Putting together the formula for the latter with that shown here, one concludes that, roughly, the requirement implies that the number of afferents of (indirect) hippocampal origin to a given neocortical stage (C^{HBP}), must be $C^{HBP} = C^{RC}a_{nc}/a_{CA3}$, where C^{RC} is the number of recurrent collaterals to any given cell in CA3, the average sparseness of a neocortical representation is a_{nc}, and a_{CA3} is the sparseness of memory representations in CA3 (Treves and Rolls, 1993).

One is led to a definite conclusion: a mechanism of the type envisaged here could not possibly rely on a set of monosynaptic CA3-to-neocortex back-projections. This would imply that, to make a sufficient number of synapses on each of the vast number of neocortical cells, each cell in CA3 has to generate a disproportionate number of synapses (i.e. C^{HBP} times the ratio between the number of neocortical cells and the number of CA3 cells). The required divergence can be kept within reasonable limits only by assuming that the back-projecting system is polysynaptic, provided that the number of cells involved grows gradually at each stage, from CA3 back to neocortical association areas (see Fig. 1).

The theory of the recall of recent memories in the neocortex from the hippocampus provides a clear view about why backprojections should be as numerous as forward projections in the cerebral cortex. The reason suggested for this is that as many representations may need to be accessed by back-projections for recall and related functions (see Rolls, 1989a,b,c) in the population of cortical pyramidal cells as can be accessed by the forward projections, and this limit is given by the number of inputs onto a pyramidal cell (and the sparseness of the representation), irrespective of whether the input is from a forward or a back-projection system.

Summary

We have considered how the neuronal network architecture of the hippocampus may enable it to act as an intermediate term buffer store for recent memories, and how information may be recalled from it to the cerebral cortex using modified synapses in back-projection pathways from the hippocampus to the cerebral cortex. The recalled information in the cerebral neocortex could then be used by the neocortex in the formation of long-term memories, which is severely impaired by damage to the hippocampus.

Acknowledgements

Different parts of the research described here were supported by the Medical Research Council,

PG8513790, by an EEC BRAIN grant, by the MRC Oxford Research Centre in Brain and Behaviour, by the Oxford McDonnell-Pew Centre in Cognitive Neuroscience, and by a Human Frontier Science program grant.

References

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

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.

Amit, D.J. (1989) *Modelling Brain Function*. Cambridge University Press, New York.

Ishizuka, N., Weber, J. and Amaral, D.G. (1990) Organization of intrahippocampal projections originating from CA3 pyramidal cells in the rat. *J. Comp. Neurol.*, 295: 580–623.

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

Rolls, E.T. (1987) Information representation, processing and storage in the brain: analysis at the single neuron level. In: J.-P. Changeux and M. Konishi (Eds.), *The Neural and Molecular Bases of Learning*, Wiley, Chichester, pp. 503–540.

Rolls, E.T. (1989a) Functions of neuronal networks in the hippocampus and neocortex in memory. In: J.H. Byrne and W.O. Berry (Eds.), *Neural Models of Plasticity: Experimental and Theoretical Approaches*, Ch. 13, Academic Press, San Diego, pp. 240–265.

Rolls, E.T. (1989b) The representation and storage of information in neuronal networks in the primate cerebral cortex and hippocampus. In: R. Durbin, C. Miall and G. Mitchison (Eds.), *The Computing Neuron*, Ch. 8, Addison-Wesley, Wokingham, England, pp. 125–159.

Rolls, E.T. (1989c) Functions of neuronal networks in the hippocampus and cerebral cortex in memory. In: R.M.J. Cotterill (Ed.), *Models of Brain Function*, Cambridge University Press, Cambridge, pp. 15–33.

Rolls, E.T. (1990a) Theoretical and neurophysiological analysis of the functions of the primate hippocampus in memory. *Cold Spring Harbor Symp. Quant. Biol.*, 55: 995–1006.

Rolls, E.T. (1990b) Functions of the primate hippocampus in spatial processing and memory. In: D.S. Olton and R.P. Kesner (Eds.), *Neurobiology of Comparative Cognition*, Ch. 12, Lawrence Erlbaum, Hillsdale, NJ, pp. 339–362.

Rolls, E.T. (1991) Functions of the primate hippocampus in spatial and non-spatial memory. *Hippocampus*, 1: 258–261.

Rolls, E.T. (1992) Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Phil. Trans. R. Soc.*, 335: 11–21.

Rolls, E.T. and O'Mara, S. (1993) Neurophysiological and theoretical analysis of how the hippocampus functions in memory. In: T. Ono, L.R. Squire, M. Raichle, D. Perrett and M. Fukuda (Eds.), *Brain Mechanisms of Perception: From Neuron to Behavior*, Oxford University Press, New York pp. 276–300.

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

Tovee, M.J., Rolls, E.T., Treves, A. and Bellis, R.P. (1993) Information encoding and the responses of single neurons in the primate temporal visual cortex. *J. Neurophysiol.*, 70: 640–654.

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

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

Treves, A. 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. and Rolls, E.T. (1994) A computational analysis of the role of the hippocampus in memory. *Hippocampus*, in press.

Treves, A., Rolls, E.T. and Tovee, M.J. (1994) On the time required for recurrent processing in the brain. *Proc. Natl. Acad. Sci. USA*, in press.