Pattern separation, completion, and categorisation in the hippocampus and neocortex

Edmund T. Rolls *

Oxford Centre for Computational Neuroscience, Oxford, England, United Kingdom
University of Warwick, Department of Computer Science, Coventry CV4 7AL, England, United Kingdom

ABSTRACT

The mechanisms for pattern completion and pattern separation are described in the context of a theory of hippocampal function in which the hippocampal CA3 system operates as a single attractor or autoassociation network to enable rapid, one-trial, associations between any spatial location (place in rodents, or spatial view in primates) and an object or reward, and to provide for completion of the whole memory during recall from any part. The factors important in the pattern completion in CA3 and also a large number of independent memories stored in CA3 include: a sparse distributed representation, representations that are independent due to the randomizing effect of the mossy fibres, heterosynaptic long-term depression as well as long-term potentiation in the recurrent collateral synapses, and diluted connectivity to minimize the number of multiple synapses between any pair of CA3 neurons which otherwise distort the basins of attraction. Recall of information from CA3 is implemented by the entorhinal cortex perforant path synapses to CA3 cells, which in acting as a pattern associator allow some pattern generalization. Pattern separation is performed in the dentate granule cells using competitive learning to convert grid-like entorhinal cortex firing to place-like fields, and in the dentate to CA3 connections that have diluted connectivity. Recall to the neocortex is achieved by a reverse hierarchical series of pattern association networks implemented by the hippocampo-cortical backprojections, each one of which performs some pattern generalization, to retrieve a complete pattern of cortical firing in higher-order cortical areas. New results on competitive networks show which factors contribute to their ability to perform pattern separation, pattern clustering, and pattern categorisation, and how these apply in different hippocampal and neocortical systems.

1. Pattern separation and pattern completion in the hippocampal system

1.1. Introduction

There is great interest in how pattern separation and pattern completion in the hippocampus contribute to its functions in memory and spatial function (Giocomo, Moser, & Moser, 2011; Hasselmo & Wyble, 1997; Hunsaker & Kesner, 2008, 2013; Jezek, Henriksen, Treves, Moser, & Moser, 2011; Kesner, 2007, 2013; Kesner, Morris, & Weeden, 2012; Kesner & Rolls, 2015; Leutgeb, Leutgeb, Moser, & Moser, 2007; McHugh et al., 2007; Nakashiba et al., 2012; Nakazawa et al., 2002, 2003; Rolls & Kesner, 2006; Rolls & Treves, 1998; Wills, Lever, Cacucci, Burgess, & O'Keefe, 2005).

The aim of Section 1 of this paper is to describe some of the different types of pattern separation and pattern completion in the hippocampal system, and the mechanisms that implement them, in part as an Introduction to the special issue of the *Neurobiology of Learning and Memory* (2016) on pattern separation and pattern completion. It is important to appreciate that there are different mechanisms each of which contributes to pattern separation or pattern completion in the hippocampal system, for this helps not only in the understanding of how the hippocampal system operates, but also helps in the evaluation of the effects of changes that influence each of these mechanisms. These different mechanisms are separated into different subsections of this paper, so that the operation and contributions of each mechanism can be clarified and evaluated. The different mechanisms for pattern separation and pattern completion are considered in the context of a theory of hippocampal function (Rolls, 2008, 2010b). More comprehensive

* Address: Oxford Centre for Computational Neuroscience, Oxford, England, United Kingdom.
E-mail address: Edmund.Rolls@oxcns.org
URL: http://www.oxcns.org

descriptions of this theory of hippocampal function, and of differences between the primate and rodent hippocampal neuronal representations and the implications for understanding human memory, are provided elsewhere (Kesner & Rolls, 2015; Rolls, 2008, 2010b, 2013; Rolls & Kesner, 2006; Rolls & Xiang, 2006). The theory presented here (Rolls' theory of hippocampal function) has been developed through many stages (Kesner & Rolls, 2015; Rolls, 1987, 1989a, 1989b, 1989d, 1990a, 1990b, 1991, 1995, 1996b, 2008, 2010b; Rolls & Deco, 2010; Rolls & Kesner, 2006; Rolls & Treves, 1998; Treves & Rolls, 1991, 1992, 1994), has as a predecessor developments made by David Marr (Marr, 1971) in understanding cortical structure and computation (though he never identified the CA3 system as an autoassociation network), and has benefitted greatly from collaborations with many whose names appear below in the citations, including Alessandro Treves and Simon Stringer.[1] The operation of pattern association networks (also known as heteroassociation networks because one pattern is associated with another pattern), autoassociation networks, and competitive networks has been described elsewhere (Hertz, Krogh, & Palmer, 1991; Rolls, 2008, 2016; Rolls & Treves, 1998).

In Section 2 of this paper, I address how, somewhat in contrast, important computational functions of the cerebral neocortex involve pattern categorisation, in which similar patterns are treated as being of the same object, and different patterns are treated as being of different objects.

In Section 3 of this paper, new and original computational analyses are described of these different processes of pattern separation, pattern completion, and pattern categorisation, which are fundamental to understanding the operation of the neocortex as well as hippocampal cortex (Rolls, 2016). These new results are shown in Figs. 5–8.

---

[1] Historical note: The theory was originally developed as described next. Marr (1971) showed how a network with recurrent collaterals could complete a memory using a partial retrieval cue, and how sparse representations could increase the number of memories stored. Marr did not apply this to particular networks within the hippocampus, and dealt with binary neurons and binary synapses which utilised heavily the properties of the binomial distribution. Treves and Rolls (1994) and Rolls and Treves (1998) have argued that approaches to neurocomputation which base their calculations on what would happen in the tail of an exponential, Poisson, or binomial distribution are very fragile, and used different techniques from theoretical physics. Early work of Gardner-Medwin (1976) showed how progressive recall could operate in a network of binary neurons with binary synapses. The analysis of these autoassociation or attractor networks was developed by Kohonen (1977, 1984), Kohonen, Oja, and Lehtio (1981) and Hopfield (1982), and the value of sparse representations was quantified by Treves and Rolls (1991). Marr's model was assessed by Willshaw and Buckingham (1990), who found that the hippocampal model described by Marr was not dependent on a recurrent collateral effect. Rolls presented a theory of the hippocampus to the Dahlem conference in 1985 on the Neural and Molecular Bases of Learning in which the CA3 neurons operated as an autoassociation memory to store episodic memories including object and place memories, and the dentate granule cells operated as a preprocessing stage for this by performing pattern separation so that the mossy fibres could act to set up different representations for each memory to be stored in the CA3 cells. He suggested that the CA1 cells operate as a recoder for the information recalled from the CA3 cells to a partial memory cue, so that the recalled information would be represented more efficiently to enable recall, via the backprojection synapses, of activity in the neocortical areas similar to that which had been present during the original episode. This appeared in print as (Rolls, 1987), and further details were made explicit by Rolls (1989a, 1989b, 1989c, 1989d). Marr (1971) did not specify the functions of the dentate granule cells vs the CA3 cells vs the CA1 cells (which were addressed in the Rolls (1989a, 1989b, 1989c, 1989d) papers and quantitatively by Treves and Rolls (1992, 1994)), nor how retrieval to the neocortex of hippocampal memories could be produced, for which a theory was developed by Rolls (1987, 1989a, 1989b, 1989d) and made quantitative by Treves and Rolls (1994). McNaughton and Morris (1987) suggested that the CA3 network might be an autoassociation network, and that the mossy fibre to CA3 connections might implement 'detonator' synapses. However, the concepts that the diluted mossy fibre connectivity might implement selection of a new random set of CA3 cells for each new memory, and that a direct perforant path input to CA3 was needed to initiate retrieval, were introduced by Treves and Rolls (1992). Contributions by Levy (e.g. (1989)), McNaughton (1991); Hasselmo; Lisman; McClelland, McNaughton and O'Reilly (McClelland, McNaughton, & O'Reilly, 1995), and many others, are described below.

A subtheme of this paper is that diluted connectivity in all the types of network described here, pattern association, autoassociation, and competitive networks, provides an important component in how they operate in the brain.

1.2. Background to the approach to hippocampal function

1.2.1. Event and episodic memory

The focus is on a fundamental property of episodic memory, the ability to store and retrieve the memory of a particular single event involving an association between items such as the place and the object or reward seen at that place. Episodic memory in the sense of a series of linked events requires this type of event memory, and could be implemented by linking together a series of events.

An event consists of a set of items that occur together, such as seeing a particular object or person's face in a particular place. An everyday example might be remembering where one was for dinner, who was present, what was eaten, what was discussed, and the time at which it occurred. The spatial context is almost always an important part of an episodic memory (Dere, Easton, Nadel, & Huston, 2008), and it may be partly for this reason that episodic memory is linked to the functions of the hippocampal system, which is involved in spatial processing and memory. The ability to recall a whole memory from a partial cue is an important property of episodic memory, and is referred to as completion.

1.2.2. Systems-level functions and connections of the primate hippocampus

Any theory of the hippocampus must state at the systems level what is computed by the hippocampus. Some of the relevant evidence about the functions of the hippocampus in memory comes from the effects of damage to the hippocampus, the responses of neurons in the hippocampus during behaviour, and the systems-level connections of the hippocampus, described in more detail elsewhere (Kesner & Rolls, 2015; Rolls, 2008, 2010b; Rolls & Kesner, 2006; Rolls & Xiang, 2006). Many of the memory functions are important in event or episodic memory, in which the ability to remember what happened where on typically a single occasion (or trial in a learning experiment) is important. It is suggested that an autoassociation memory implemented by the CA3 neurons enables event or episodic memories to be formed by enabling associations to be formed between spatial and other including object or reward representations, and for completion to then occur in recall. An important property of this autoassociation and completion is that completion of a whole memory can occur from any part. This is different from pattern association memory, in which a visual stimulus might become associated with a taste by associative synaptic modification. Later presentation of the visual stimulus would retrieve the taste representation. However, presentation of the taste would not retrieve the visual representation, and this is an important and fundamental difference between autoassociation and pattern association, as described in detail elsewhere (Rolls, 2008, 2014; Rolls & Treves, 1998).

Information stored in the hippocampus will need to be retrieved and affect other parts of the brain in order to be used. The information about episodic events recalled from the hippocampus could be used to help form semantic memories (Rolls, 1989b, 1989d, 1990a; Treves & Rolls, 1994). For example, remembering many particular journeys could help to build a geographic cognitive map in the neocortex. The hippocampus and neocortex would thus be complementary memory systems, with the hippocampus being used for rapid, "on the fly", unstructured storage of information involving activity potentially arriving from many areas of the neocortex; while the neocortex would gradually build and adjust on the basis of much accumulating information, often recalled from the hippocampal unstructured store, the semantic