THE HIPPOCAMPUS AS A COGNITIVE MAP

JOHN O'KEEFE
AND
LYNN NADEL

1978
CLARENDON PRESS · OXFORD

Oxford University Press, Walton Street, Oxford OX2 6DP

OXFORD LONDON GLASGOW NEW YORK
TORONTO MELBOURNE WELLINGTON CAPE TOWN
IBADAN NAIROBI DAR ES SALAAM LUSAKA ADDIS ABABA
KUALA LUMPUR SINGAPORE JAKARTA HONG KONG TOKYO
DELHI BOMBAY CALCUTTA MADRAS KARACHI

British Library Cataloguing in Publication Data
O'Keefe, John
The hippocampus as a cognitive map.
I. Hippocampus (Brain)
I. Title II. Nadel, Lynn
612´.825 QP381 77-30572

ISBN 0-19-857206-9

Filmset by Northumberland Press, Gateshead, Tyne and Wear
Printed in Great Britain
by Richard Clay (The Chaucer Press) Ltd., Bungay, Suffolk

TO

E. C. TOLMAN
Who first dreamed of cognitive maps in rats and men

D. O. HEBB
Who taught us to look for those maps in the brain

AND

A. BLACK
Who insisted that we pursue our route with rigour

Preface

Scientific theories have been likened to maps.* Like maps they provide a means for finding one's way in an unknown domain. And also like maps, they are a social product, the culmination of considerable effort by a large number of people. The cartographer draws on the lessons of past geographers, he builds on the work of the explorer and the surveyor, he relies on the skills of the geometer no less than those of the draughtsman and the parchment-maker. In this preface, we can do little more than pay small tribute to some of those who have contributed to the making of this map.

Our voyage of discovery was launched from the McGill Psychology Department where we were both Ph.D. students in the 1960s. At McGill, Don Hebb had built a department which encouraged students to theorize about the neural bases of perception, motivation, and cognition, and which gave them the freedom and the opportunity to test their ideas. Hebb's theory of the cell assembly was like a model map–it showed what a theory of the brain would look like. Hebb's graduate seminar was particularly influential. He emphasized the extent to which behaviourist notions could handle most of the available data and pointed to those few areas which showed the limitations of this approach: latent learning, sensory preconditioning, the surprise of Tinklepaugh's monkeys. At McGill one learned techniques from post-doctoral fellows and other graduate students. We were particularly fortunate to be able to learn from Ken Casey, now at Michigan, and Herman Bouma, now at the Institute for Perception, Eindhoven. Our research directors Ron Melzack and Dalbir Bindra were generous and tolerant to a fault.

When we first started to explore this terra incognita, we found only a small number of footprints in the sand. Most prominent were those of Helen Mahut, Case Vanderwolf, and Jim Ranck. We were amongst friends: both Helen and Case were McGill graduates and Jim had been a colleague of Ken Casey at Michigan for several years. To be sure others had also sailed past this coast; but, having dimly sighted it from afar, they quickly pushed on towards those ever-beckoning chimera: the Quagmire of Inhibition, the Palisades of Persistence, and the Channel of Attention. Helen Mahut was one who landed, planted her feet and stayed. Her demonstration of a selective deficit in spatial reversal learning after damage

* This analogy was first suggested by Toulmin (1953). In Chapter 2 we elaborate on this distinction between maps and routes. Our extension of his analogy here to the social aspects of science derives to some extent from conversations with Graham Goddard.

to the hippocampal system was an important pointer to the spatial function of the hippocampus.

Another explorer who signposted the way was Case Vanderwolf. His observation that hippocampal theta in the rat was in some way bound up with a class of movements received early confirmation by Abe Black's impressive analytic experiments. Taken together, these had a strong influence on our thinking about hippocampal function. Not that we ever accepted Vanderwolf's implicit assumption that the function of the hippocampus could be equated with the behavioural or psychological correlate of the hippocampal EEG. Rather we concluded that a theory of hippocampal function had to account for the availability to the hippocampus of information about a class of movements.

Jim Ranck pioneered the recording of single cells from the hippocampus of the freely moving animal. Others had done so before him (most notably Olds (1965) and Vinogradova (1970)) but under very restricted conditions which may have prevented them from seeing the more spatial aspects of these cells. Jim Ranck's rats, like ours, were allowed to move about in an extended environment. The remarkable similarity between his data and ours was a source of great encouragment in those early days. We used different terminology and emphasized different aspects of the behavioural correlates but it was clear that we were seeing the same things. Many of our apparent discrepancies were straightened out during the three months that Jim Ranck spent in our laboratory in 1972 and during our subsequent visits to his laboratories in Michigan and New York. Some of our ideas on the neuroethological approach to single-unit recording are identical to his as set out in an unpublished paper dating from 1972.

More recently Dave Olton and Phil Best have added to our understanding of the spatial function of the hippocampus. They have confirmed the spatial nature of the hippocampal cells. Olton's eight-arm maze task appears to be a particularly sensitive test of the hippocampal mapping system.

Much of the past seven years has been occupied with the details of the map: delimiting the boundaries of the theory, knitting together the disparate pieces and patches, deriving detailed predictions about the behaviour of lesioned animals, designing experiments to test them. During this time we have had tremendous help from Abe Black. His investigations on the behavioural correlates of the hippocampal EEG and the effects of hippocampal lesions are models of experimental design. Some of the more explicit and testable predictions of the theory can be traced back to conversations with Abe Black. He has a complete collection of all of the many versions of this book, most of them heavily annotated and dog-eared. We decided it was time to publish when we found that we could actually read much of the typed text in the most recent one.

Other dog-eared, mind-chewed copies belong to Graham Goddard and Jim Ranck. Graham Goddard wrote an extensive commentary on the original version while patiently waiting in queues to get into the Tutankhamun exhibition at the British Museum. Jim Ranck read several chapters while trapped in the New York subway during a recent power failure. The theory has been tested under a wide range of environmental conditions.

Others who tried to improve the book through their comments on parts or all of it include Per Andersen, Herman Bouma, Hide Ishiguro, Peter Molnar, Eileen O'Keefe, Geoff Raisman, Tim Shallice, Pat Wall, and several anonymous advisers for the Oxford University Press. Peter Molnar read the book in proof with great care and pointed out numerous mistakes. One of the advisers took the trouble to point out many errors in the original version and prompted revisions of several sections. A more recent adviser chided us for our chiaroscuro painting of the data, but we had decided early on to present the boldest black-and-white case for the theory; later will be time enough to etch in the subtle greys.

Over the past seven years, we have been extremely fortunate in our colleagues and students, some of whom have been our strongest critics. Among those whose research is cited in this book are Abe Black, Dulcie Conway, Richard Morris, Liz Somerville, Simon Keightley, Dave Kill, and Barbara Oetliker. Our experiments were possible only because of skilled technical help we have had from Alan Ainsworth and Jitendra Patel. Jane Astafiev drew many of the figures and Julia O'Connor typed and repeatedly retyped the manuscript. Together with Howard O'Connor she also prepared the author index. Financial support has been provided by the Science Research Council, the Medical Research Council, and the Wellcome Trust, all of the United Kingdom.

Perhaps our greatest thanks are due to those who created a safe haven for us to work in during these past seven years. Maps are not made during a state of siege. Most of the book was written while we were honorary research fellows in the Cerebral Functions Group, in the Department of Anatomy, University College London. J. Z. Young, the head of the department until 1974, has had a life-long interest and involvement in the neural basis of animal memory. He advised as, encouraged us, and housed us. His hospitality has been continued by the present head, G. Burnstock.

Pat Wall, head of the Cerebral Functions Group, more than anyone else deserves credit for the existence and present shape of this book. It was he, who, in 1971, originally suggested that we turn a tentative review article into a slim monograph and then had to witness its inexorable growth year after year to reach its present girth. Throughout he has encouraged and cajoled us, defended us, read what we wrote, criticized, gently prodded, pushed, and has patiently given us the time to work things out. In these rapid-transit days of publish or perish, he slowed time to our snail's pace.

L.N. would also like to thank Holger Ursin and Bolek Srebro for their hospitality during the six months he spent at the University of Bergen and Larry Squire for his hospitality during the eleven months he spent in the Department of Psychiatry, University of California, San Diego.

The staff of the Oxford University Press skilfully steered the manuscript through the many phases of publication. We thank them for their patience, help, and good humour.

Friends and family put up with us all these years, successfully keeping us out of the asylum and restraining themselves from committing homicide: Eileen, Kieron, Riley, Maria, Britt, Melissa, Kenny, Mischa, Joao, Diane, Garth, Fiz, Wanda, and Charlie. ‘Thank you’ is small recompense.

London J.O'K.
June 1978 L.N.

Acknowledgements

WE should like to thank the following scientists for permission to use material from their publications: P. Andersen, A. Arduini, R. B. Chronister, A. H. Lohman, J. B. Ranck, L. J. Stensaas, C. H. Vanderwolf, G. Van Hoesen, I. Q. Whishaw, and J. Winson.
The following organizations have kindly allowed us to reproduce material: Academic Press, New York; the American Association for the Advancement of Science, Washington, D.C.; the American Physiological Society, Bethesda; the American Psychological Association, Washington, D.C.; Johann Ambrosius Barth, Leipzig; J. F. Bergmann-Verlag, Munich; Consejo Superior de Investigaciones Cientificos, Madrid; Elsevier/North Holland Biomedical Press, Amsterdam; S. Karger A.G., Basel; London Transport; the Ordnance Survey; H.M. Stationery Office; Plenum Press, New York; Wistar Press, Philadelphia.

Contents

Introduction

THIS book is concerned with three topics which, at first glance, do not appear to be related: (1) a part of the brain known as the hippocampus; (2) the psychological representation of space; (3) context-dependent memory. We shall argue that the hippocampus is the core of a neural memory system providing an objective spatial framework within which the items and events of an organism's experience are located and interrelated. In this introduction we shall briefly summarize the main points of our argument.

The first chapter begins with an analysis of the history of ideas about space, as developed primarily by philosophers and influenced by physicists and mathematicians. Here, our primary concern will be to show that traditionally there have been two incompatible theories of the nature of physical space. The *absolute* theory views space as a stationary framework which is separate from, but can nevertheless contain, material objects; the *relative,* or relational,* theory views space as nothing more than the relations between material objects themselves. These arguments, which relate purely to an understanding of the physical universe, have had a profound effect on thinking about ways in which organisms represent space. In particular, the advent of non-Euclidean geometries and the theory of relativity have shifted the argument dramatically in the direction of a relative theory of both physical and psychological space, with the consequence that most authors attempt to derive all psychological notions of space from an organism's interaction with objects and their relations. The notion of an absolute spatial framework, if it exists at all, is held by these authors to derive from prior concepts of relative space, built up in the course of an organism's interaction with objects or with sensations correlated with objects.

In contrast to this view, we think that the concept of absolute space is primary and that its elaboration does not depend upon prior notions of relative space. In our view, organisms represent space in several independent, though interrelated, ways. A number of neural mechanisms generate psychological spaces referred to the observer, and these are consistent with a relative theory of space. Amongst these are spaces centred on the eye, the head, and the body, all of which can be subsumed under the heading of *egocentric space.* In addition, there exists at least one neural system which provides the basis for an integrated model of the environment. This system underlies the notion of absolute, unitary space, which is

* The terms relative and relational have been used interchangeably by most authors. While Lacey (1970) argues that there is a distinction to be drawn between the two, for our purposes this distinction is unimportant.

a non-centred stationary framework through which the organism and its egocentric spaces move.

We shall call the system which generates this absolute space a *cognitive map* and will identify it with the hippocampus.* The term was first used by Tolman (1948), and has recently reappeared in the work of some psychologists and geographers interested in the perception of large-scale environments (e.g. Menzel 1973, Richards 1974). Although this use of the concept of cognitive mapping has descriptive and predictive value, it is too vague for the present book. To serve our purpose adequately it must be specified in sufficient detail to enable us to translate it into a neural model.

Briefly, a cognitive map would consist of two major systems, a *place* system and a *misplace* system. The first is a memory system which contains information about places in the organism's environment, their spatial relations, and the existence of specific objects in specific places.** The second, misplace, system signals changes in a particular place, involving either the presence of a new object or the absence of an old one. The place system permits an animal to locate itself in a familiar environment without reference to any specific sensory input, to go from one place to another independent of particular inputs (cues) or outputs (responses), and to link together conceptually parts of an environment which have never been experienced at the same time. The misplace system is primarily responsible for *exploration,* a species-typical behaviour which functions to build maps of new environments and to incorporate new information into existing maps.

In order to set this cognitive mapping system within the more general context of brain and behaviour we shall find it necessary to describe briefly a model of animal learning which lays strong emphasis on the use of *hypotheses* (cf. Tolman and Krechevsky 1933). Such a model states that, when faced with a problem, animals do not respond randomly, but rather select and test hypotheses, or strategies, concerning the solution. We assume that one form of hypothesis involves approaching or avoiding places in the environment and, further, that this type of hypothesis is based on information contained within the hippocampal cognitive mapping system. Other hypotheses, such as those involving an approach to a specific object or the performance of a particular response, are based on information stored by other neural systems. While our discussion will touch upon the properties of all these forms of hypothesis, our emphasis will remain on place hypotheses and the cognitive map.

* Strictly speaking, the hippocampus should be called a cognitive mapping system, and the term cognitive map reserved for the products of that system. For convenience, we shall use the latter term to refer to both the neural structure and the representation of a particular environment; the context should make our meaning clear.

** Simply stated, a place is a part of absolute space and can only be defined in terms of its locus within the neural system mediating that concept. We shall say more about this central concept later (see p. 93).

On the basis of a review of the anatomy and physiology of the hippocampus a structural model of the cognitive map will be proposed. Amongst other things, this model will draw upon the existence of place-coded neurones in the hippocampus and will also propose specific functions for the rhythmic electrical activity *(theta)* recorded from the hippocampus during certain behaviours. In conjunction with the general behavioural model specified earlier, this structural model will enable us to generate predictions about the changes in behaviour brought about by dysfunction of the hippocampus. Here, we shall attempt to review the available literature from experiments with a variety of species, and show that it is consistent with the theory. The work with infra-humans will be discussed first, followed by an analysis of the relevant human clinical data. It will be suggested that the left hippocampus in humans functions in *semantic mapping,* while the right hippocampus retains the spatial mapping function seen in infra-humans. On this view, species differences in hippocampal function reflect changes in the inputs to the mapping system, rather than major changes in its mode of operation. This proposal offers the hope of unifying animal experimental and clinical approaches to the problem of hippocampal function.