Parallel Distributed Processing

Implications for Psychology and Neurobiology

Edited by
R. G. M. MORRIS

OXFORD SCIENCE PUBLICATIONS

Parallel Distributed Processing

Implications for Psychology and Neurobiology

Edited by
R. G. M. MORRIS
University of Edinburgh

CLARENDON PRESS · OXFORD
1989

Oxford University Press, Walton Street, Oxford OX2 6DP
Oxford New York Toronto
Delhi Bombay Calcutta Madras Karachi
Petaling Jaya Singapore Hong Kong Tokyo
Nairobi Dar es Salaam Cape Town
Melbourne Auckland
and associated companies in
Berlin Ibadan

Oxford is a trade mark of Oxford University Press

Published in the United States
by Oxford University Press, New York

British Library Cataloguing in Publication Data
Parallel distributed processing: implications for
psychology and neurobiology.
1. Man. Cognition. Neuropsychological aspects
I. Morris, R. G. M.
153.4
ISBN 0-19-852178-2

Library of Congress Cataloging in Publication Data
Parallel distributed processing: implications for psychology and neurobiology/edited by R. G. M. Morris.
Based on a conference organized by the Experimental Psychology Society and held at the University of Oxford on July 1, 1987.
Includes bibliographies and indexes.
1. Cognition—Congresses. 2. Parallel processing (Electronic computers)—Congresses. 3. Electronic data processing—Distributed processing—Congresses. 4. Neural circuitry—Congresses. 5 Neurobiology—Congresses. 6. Human information processing—Congresses. I. Morris, R. G. M. (Richard G. M.) II. Experimental Psychology Society.
[DNLM: 1. Cognition—congresses. 2. Mental Processes—congresses. 3. Models, Psychological—congresses. 4. Nervous System—physiology—congresses. WL 102 P2215 1987]
BF311.P3133 1989
153—dc20
89-8699
ISBN 0-19-852178-2

Typeset by
Cotswold Typesetting Limited, Gloucester
Printed in Great Britain by
Alden Press Ltd, Oxford

Parallel Distributed Processing

Preface

Recent years have witnessed a substantial growth of interest in 'Parallel distributed processing', 'connectionist', or 'neural network' models of cognitive function. Such models have been in existence for many years, but the publication of Hinton and Anderson's (1981) book *Parallel models of associative memory* (Lawrence Erlbaum Associates), followed, in 1986, by the two-volume book by Rumelhart, McClelland, and their colleagues entitled *Parallel distributed processing: explorations in the microstructure of cognition* (Bradford Books, MIT Press) did much to attract the attention of experimental psychologists. Beginning in 1987, PDP 'workshops' sprang up in several British and North American Universities, with the participants working their way through the chapters in these books and discussing their implications.

Throughout the 1970s and 1980s, many neurobiologists also developed an interest in trying to understand how complex networks of real neurons perform various tasks. The early papers of David Marr, discussing the cerebellum, archicortex and neocortex, are perhaps the best-known examples of this approach (e.g. 'Simple memory: a theory for archicortex', *Philosophical Transactions of the Royal Society*, 1971, **262,** 23–81). However, in his 1982 book *Vision*, Marr expressed some intellectual disappointment with this early work, worrying that these models failed to grasp the complexities of the algorithms being computed by complex networks. In its place, he outlined an approach which, while emphasizing the importance of different levels of explanation for the neurosciences, also stressed the need for these different levels to be bridged. Sadly, David Marr did not live to see the promise of his new approach fulfilled. Many neurobiologists now hope that current developments in the formal analysis of neural networks will provide just such a bridge between psychological accounts of cognitive function and accounts couched at the level of real neurons. In addition, new tract-tracing, immunocytochemical and recording techniques make it possible to describe the detailed course and topography of neural interconnections within discrete networks, the neurotransmitters used, and the capacities of specific pathways for synaptic plasticity in sufficient detail to make modelling worthwhile.

The central aim of the present volume is to ask the question: 'What are the implications of these new parallel distributed processing (PDP) models?' Or, to put it another way, should those experimental psychologists and neurobiologists interested in cognitive function set about their experiments differently in the light of these developments? Clearly, this is an issue on which there is a tremendous difference of opinion. Some see the PDP approach as representing a great step forward in the effort to build neurally realistic models of sufficient sophistication to capture the detailed microstructure of cognition. Others worry that the explanations offered are deceptive. In particular, some neurobiologists worry that just because neurons are arranged in large parallel networks is no reason in itself to suppose that they are carrying out their processing using algorithms like back-propagation which are presently the focus of so much attention as models of cognitive function.

This book has emerged out of a one-day conference organized by the Experimental Psychology Society and held at the University of Oxford on 1 July 1987. That meeting was organized into a morning session devoted to work on human perception, memory, and language function, and an afternoon session devoted to psychological and neurobiological work on animals. The organization of the book is slightly different, partly to emphasize the different levels of explanation characterized by the psychological and neurobiological approach. There are three sections. Part I is concerned with *formal models.* It introduces the approach and discusses the all-important assumptions and algorithms of PDP models. Part II is concerned with *implications for psychology* and covers both human and animal research. This section describes the attempts of three different groups of experimenters to consider the relevance of PDP-type models, but also includes one chapter critical of the approach. Part III is concerned with *implications for neurobiology.* Here also, three authors are impressed by the force of the neural network approach, one is more cautious. Each of the sections of the book is introduced by a short chapter sketching out some of the issues discussed and alluded to in the main chapters that follow. Some readers may find it helpful to read these three chapters first to get an idea of the scope of the book.

Organizing the meeting in Oxford and putting this book together has been both a privilege and a pleasure, but are two tasks which would not have been possible without the help of others. I am particularly grateful to Drs Brian Rogers and Peter McLeod, both of the Department of Experimental Psychology in Oxford, who made all the necessary preparations, including a television relay into an adjoining lecture theatre—such was the interest in the meeting. I am also grateful to Professor L. Weiskrantz for his permission to use the facilities of his Department. Professor N. S. Sutherland and Dr David Willshaw acted as Chairmen for the morning and afternoon sessions and kept things running smoothly and to time. The contributors to the book tolerated and, in all cases, accepted my requests for minor changes in their manuscripts for the sake of continuity and consistency in the book. They also kindly agreed to offer their support to a fund recently set up by the Experimental Psychology Society to support research by postgraduate students. Two of the chapters have been published previously. Geoffrey Hinton's chapter has previously appeared in a volume published by Lawrence Erlbaum; while Steven Pinker's chapter, originally written for the present book, has also appeared in *Trends in*

Neuroscience. I am grateful to Erlbaum Associates and Elsevier for allowing the present reprinting. Finally, I am grateful to Leslie Chapman for her assistance with the Index and to the staff of the Oxford University Press who have supported the project from the outset and who have so ably seen it through to completion.

University of Edinburgh R.G.M.M.
October 1988

Contributors

Richard Granger Centre for the Neurobiology of Learning and Memory, University of California, Irvine, California, 92717 USA
Robert D. Hawkins Centre for Neurobiology and Behaviour, Columbia University, 722 West 168th Street, New York, New York 10032 USA
Geoffrey E. Hinton The Canadian Institute for Advanced Research and Departments of Computer Science and Psychology, University of Toronto, Toronto M5S 1A4, Canada
Glyn W. Humphreys Cognitive Science Research Centre, School of Psychology, University of Birmingham, PO Box 363, Birmingham B15, England
Helen Kaye Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge, CB2 3EB, England
H. Christopher Longuet-Higgins, F.R.S. Laboratory of Experimental Psychology, University of Sussex, Brighton, BN1 9QG, England
Gary S. Lynch Centre for the Neurobiology of Learning and Memory, University of California, Irvine, California, 92717, USA
Nicholas J. Mackintosh, F.R.S. Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge, CB2 3EB, England
Ian P. L. McLaren Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge, CB2 3EB, England
James L. McClelland Department of Psychology, Carnegie-Mellon University, Pittsburgh, Pennsylvania, 15213 USA
Richard G. M. Morris Department of Pharmacology, University of Edinburgh Medical School, 1 George Square, Edinburgh, EH8 9JZ, Scotland
Karalyn Patterson MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge, CB2 3EF, England
Steven Pinker Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
Alan Prince Program in Linguistics and Cognitive Science, Brandeis University, 415 South Street, Waltham, Massachusetts 02154, USA
Edmund T. Rolls Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford, OX1 3UD, England
Mark S. Seidenberg Department of Psychology, McGill University, Montreal, H3A 2T5, Canada
Bruce A. Whittlesea Department of Psychology, Mount-Allison University, Sackville, New Brunswick, E0A 3C0, Canada

Contents

Part I

Formal models

1

Network models of the mind

H. CHRISTOPHER LONGUET-HIGGINS

The association of ideas

The fact that people can associate ideas, such as names and faces, or sounds and symbols, is too obvious to need documentation. If there were a limited number of possible ideas, the developing brain could allocate a separate neurone to each, and connect every pair of such neurones by a modifiable synapses, to be facilitated if and only if the two ideas occurred in association. Ideas can, however, be very complicated, so the number of possible ideas is enormous—far too large for each to be assigned a neurone on the off-chance of it turning up. Complex ideas, 'patterns', must therefore be indexed by the association of simpler ones, 'features', to which neurones can be allocated without extravagance. By recording the pairwise associations between the features of a pattern we can set up a simple associative memory (Marr 1969), from which the pattern can be recovered by activating enough of its features and allowing these to activate the remainder. Memories of this kind not only solve the problems of 'store allocation' and 'content-addressability', but are also relatively robust against partial corruption of the contents.

Becoming more ambitious, we may attempt to use an associative net for storing not just one but several patterns (Willshaw, Buneman, and Longuet-Higgins 1969). If two such patterns share any features, there is a possibility of 'cross-talk' between them when either pattern is retrieved. This can be regarded either as a nuisance or as a bonus, according to intellectual taste: the convergent thinker will see it as a limitation on the accuracy of recall, the divergent as a source of creative generalization. But in its powers of generalization the associative net is subject to the same sort of limitations as the one-layer perceptron (Rosenblatt 1962), since it is essentially a battery of such perceptrons working in parallel on the same set of input elements. Not that this is a crippling disability; associative nets can learn to reason inductively (Willshaw 1972) and to supply the best completion of any pattern picked from an ensemble in which the accessible elements of a pattern supply *independent* clues about the inaccessible elements (Minsky and Papert 1969; Hinton and Sejnowski 1983). But a number of vital cognitive skills such as concept formation and language acquisition are known to lie beyond the competence of the associative net (Hinton and Anderson 1981), so the question arises whether all parallel distributed processing (PDP) networks of this general type are subject to similar limitations.

Two different questions

There are two questions to be asked about a PDP network or, indeed, any model of mental or cerebral activity:

1. What tasks can such a network be designed to perform?
2. Which of these tasks can the network learn to perform?

The questions must be distinguished, since most machines, however well they perform their intended functions, cannot learn anything at all. For the perceptron (i.e. the one-layer perceptron, unless otherwise stated) and the associative net, ancestors of the PDP network, the answers to both questions are known.

The perceptron is, in essence, a device for dividing bit-patterns into two classes. Each element of the pattern is registered by a separate unit, and these units are directly connected by lines of modifiable 'weight' to an output unit with an adjustable threshold. Whether or not the output unit fires depends on whether the sum of the weights on the lines from the 'active' units—those which register a 1 rather than a 0—does or does not exceed the output threshold. If and only if there exists some plane, in the space of possible patterns, that separates the patterns of one class from those of the other, then a suitable choice of weights and output threshold will ensure that the perceptron correctly distinguishes between patterns of the two kinds. Thus one can make a perceptron that distinguishes the 2-bit pattern (0,0) from any other 2-bit pattern, but no choice of weights and threshold will enable the patterns (0,1) and (1,0) to be distinguished from either of the other two, because there is no straight line in 2D separating the points (0,1) and (1,0) from the points (0,0) and (1,1). It is in this sense that the perceptron cannot solve the 'exclusive or' problem—that of telling when just one of the two input elements, but not both, has the value 1.

Remarkably enough, if a given task can be performed by a suitably prepared perceptron, then an unprepared perceptron of similar architecture can learn to perform the task. This result follows directly from the perceptron convergence theorem (Minsky and Papert 1969). It also holds for the associative net because, as already remarked, such a net is nothing more than a battery of perceptrons working in parallel, with every output unit directly connected to every input unit. It must be emphasized, however, that the restriction to 'linear' tasks, coupled with the proven ability to learn any such task, applies only to the one-layer perceptron and to its offspring the associative net.

For PDP networks with hidden units the two questions posed above remain largely unanswered, but one useful result is available, namely that a two-layer perceptron—one with a single layer of hidden units between the input layer and the output unit—can, if suitably prepared, compute any Boolean function of the input vector. Thus the 'exclusive or' predicate, the Waterloo of the one-layer perceptron, can easily be evaluated by a two-layer perceptron if the number of units and the weights of the various connections are suitably chosen (Rumelhart and McClelland 1986). The generalization of this result to PDP networks is the proposition that any mapping whatever between Boolean input vectors and output vectors can be achieved by a network with a single layer of hidden units. Unfortunately, neither result is of much practical importance, because, for an arbitrary Boolean function and a sizeable number of input units, the number of hidden units required would be quite absurdly large.

As yet, no result equivalent to the perceptron convergence theorem has been established for the multilayer perceptron, but the recently invented 'back-propagation' procedure of Rumelhart, Hinton, and Williams (1986) is a natural generalization of the perceptron learning algorithm. It has been applied with impressive results to a number of learning tasks, and in the other two chapters of this section two more such tasks crumble beneath the same steam roller.

Glimmerings of intelligence

In the good old days of the von Neumann computer (the one on your desk) one could either theorize about the correctness of programs or actually write programs to impress the onlooker with the wonders of artificial intelligence. What should have been the program to end all such programs was the general problem solver, GPS for short, of Newell, Shaw, and Simon. It used to remind one of the patent beetle killer consisting of two wooden blocks with the directions: 'Place beetle on block A and strike smartly with block B'. Perhaps there is a message here for anybody who expects PDP networks to solve all our computing problems: the representation of the problem, the choice of architecture for the network and the control of its activity may well be the most challenging parts of the enterprise.

The chapters by Hinton and by McClelland are, in their separate ways, pioneering studies in PDP modelling. It is worth reflecting why an associative net without hidden units could never learn the two family trees that Hinton's network appears to master. It would, for example, be unable to 'notice' (becuase of its incompetence with the 'exclusive or') that person 1 and person 2 always belong to the same subset of the individuals mentioned (the English family or the Italian family), never to alternative subsets. Hinton is well aware that a considerable element of intuition is involved in the design of a PDP network to solve a given learning task; but as he points out, rule-governed network design is unlikely to become a reality until we have a good mathematical theory of learning tasks in general.

McClelland's chapter gives substance to Piaget's notion of 'the equilibration of structures' in the mind of the developing child. The scientific status of this concept has always seemed a little precarious: how could one submit it to logical or experimental test? McClelland's PDP model for Siegler's balance beam task meets the case admirably, and interprets in detail a number of striking facts about the stages through which children pass in learning the task. Were that all, one might feel inclined merely to add it to the growing list of successful PDP models; but McClelland has succeeded in distilling from the back-propagation algorithm used by the model a learning principle that he states in the following terms:

> Adjust the parameters of the mind in proportion to the extent to which their adjustment can produce a reduction in the discrepancy between expected and observed events.

Such adjustments are, as he points out, exactly what are called for by the back-propagation algorithm, which Hinton explains in his chapter and also uses in his relationship-learning network.

Where we are

Everyone seems to agree that we would dearly like to have more theorems about what can or cannot be learned by PDP networks, and what architectures are required for the acquisition of given sorts of skill. In the past the computational modelling of cognitive skills (Longuet-Higgins 1987) has been carried out in languages designed for serial rather than parallel computers, but such work is not necessarily outdated by a shift of emphasis in the direction of PDP. The issue is, in any case, less a matter of principle than of implementation. In the meantime it is surely to be hoped that the art of PDP modelling will soon mature into a computational technique at least as reliable and versatile as more conventional methods of cognitive modelling.

Acknowledgements

My thanks are due to Stuart Sutherland and David Willshaw for useful comments, and to the Royal Society and the Science and Engineering Research Council for research support.

References

Hinton, G. E. and Anderson, J. A. (eds.) (1981). *Parallel models of associative memory*. Erlbaum, Hillsdale, N.J.

Hinton, G. E. and Sejnowski, T. J. (1983). Optimal perceptual inference. *Proceedings of the IEEE Conference on Computer Vision Pattern Recognition*, 448–53. IEEE Computer Society Press, Silver Spring, Maryland.

Longuet-Higgins, H. C. (1987). *Mental processes: studies in cognitive science*. MIT Press, Cambridge, Mass.

Marr, D. (1969). A theory of cerebellar cortex. *Journal of Physiology* (*Lond*) **202,** 437.

Minsky, M. and Papert, S. (1969). *Perceptrons*. MIT Press, Cambridge, Mass.

Newell, A., Shaw, J. C. and Simon, H. A. (1960). A variety of intelligent learning in a general problem solver. In *Self-organizing systems*, (eds. M. C. Yovits and S. Cameron) pp. 153–189. Pergamon Press, London.

Rosenblatt, F. (1962). *Principles of neurodynamics*. Spartan, New York.

Rumelhart, D. E. and McClelland, J. L. (1986). *Parallel distributed processing*. MIT Press, Cambridge, Mass.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). *Nature* (*Lond.*) **323,** 533.

Willshaw, D., Buneman, O. P., and Longuet-Higgins, H. C. (1969). A non-holographic model of associative memory. *Nature* (*Lond.*) **222,** 960–67.

Willshaw, D. J. (1972). A simple network capable of inductive generalization. *Proceedings of the Royal Society of London B* **182,** 233–47.