COMPUTERS, BRAINS AND MINDS

Essays in Cognitive Science

Edited by

PETER SLEZAK

Centre for Cognitive Science, University of New South Wales

and

W. R. ALBURY

School of Science and Technology Studies, University of New South Wales

KLUWER ACADEMIC PUBLISHERS

DORDRECHT / BOSTON / LONDON

Library of Congress Cataloging in Publication Data

CIP

Computers, brains, and minds : essays in cognitive science / edited by Peter Slezak and W.R. Albury.
p. cm. -- (Australasian studies in history and philosophy of science ; v. 7)
Papers from a special symposium organized for the joint conference of the Australasian Association for the History, Philosophy, and Social Studies of Science and the Australasian Association of Philosophy held in late 1985 at the University of New South Wales in Sydney, Australia.
Includes index.
ISBN 9027727597
1. Cognitive psychology--Congresses. 2. Human information processing--Congresses. 3. Neuropsychology--Congresses. 4. Cognition--Congresses. I. Slezak, Peter, 1947- . II. Albury, W. R. III. Australasian Association for the History, Philosophy, and Social Studies of Science. IV. Australasian Association of Philosophy. V. Series.
BF311.C59 1988
153--dc19 88-14698
CIP

Published by Kluwer Academic Publishers,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands

Kluwer Academic Publishers incorporates
the publishing programmes of
D. Reidel, Martinus Nijhoff, Dr W. Junk and MTP Pres

Sold and distributed in the U.S.A. and Canada
by Kluwer Academic Publishers,
101 Philip Drive, Norwell, MA 02061, U.S.A.

In all other countries, sold and distributed
by Kluwer Academic Publishers Group,
P.O. Box 322, 3300 AH Dordrecht, The Netherlands

Printed in The Netherlands

COMPUTERS, BRAINS AND MINDS

AUSTRALASIAN STUDIES IN HISTORY AND PHILOSOPHY OF SCIENCE

VOLUME 7

FOREWORD

The institutionalization of History and Philosophy of Science as a distinct field of scholarly endeavour began comparatively early – though not always under that name – in the Australasian region. An initial lecturing appointment was made at the University of Melbourne immediately after the Second World War, in 1946, and other appointments followed as the subject underwent an expansion during the 1950s and 1960s similar to that which took place in other parts of the world. Today there are major Departments at the University of Melbourne, the University of New South Wales and the University of Wollongong, and smaller groups active in many other parts of Australia and in New Zealand.

"Australasian Studies in History and Philosophy of Science" aims to provide a distinctive publication outlet for Australian and New Zealand scholars working in the general area of history, philosophy and social studies of science. Each volume comprises a group of essays on a connected theme, edited by an Australian or a New Zealander with special expertise in that particular area. Papers address general issues, however, rather than local ones; parochial topics are avoided. Furthermore, though in each volume a majority of the contributors is from Australia or New Zealand, contributions from elsewhere are by no means ruled out. Quite the reverse, in fact - they are actively encouraged wherever appropriate to the balance of the volume in question.

R. W. HOME
General Editor
Australasian Studies in History and Philosophy of Science

PREFACE*

The present volume emerges from an attempt to bring together such interest in cognitive science as there was in Australia and New Zealand in late 1985. The occasion was a special symposium organised for the joint conference of the Australasian Association for History, Philosophy and Social Studies of Science and the Australasian Association of Philosophy, held at the University of New South Wales in Sydney, Australia. Remote though Australia may be physically from the principal centres of activity in this area, in terms of Dennett's logical geography,[1] it remains to be determined where *Terra Australis* should appear in its relation both to the 'East Pole' of MIT High Church Computationalism and to the 'Zen Holism' of the American West Coast. Contrary to the flat earth appearance of Dennett's map, the contributions to this volume suggest that going far enough West brings the traveller back to the East Pole, with the antipodes standing nearer to the mother church than might be thought from mere compass direction. Be that as it may, it was the first visit to our far flung shores of the reigning Pope of Computationalism, Jerry Fodor, which was the occasion for a revitalisation of the faith (though, of course, there are well-known exegetical grounds for holding that Australian philosophers were the earliest founders of the religion, in the form of 'old testament' materialism as expounded by Place, Smart and Armstrong). Thus, Fodor's visit to give the keynote address to the Australasian cognitive science symposium was a timely event, perhaps even more a 'revival' meeting than a papal appearance, since it was also the occasion for announcing the establishment of the first graduate degree program in cognitive science in Australia at the University of New South Wales.

Given the origins of this volume in the 1985 symposium, most of the papers appearing in it have an Australian connection, with some of them, like Fodor's, actually presented for the first time at the symposium itself. Others, like Michael Arbib's and Bernard Berofsky's, although not presented at the symposium, were contributed at the invitation of the editors. Though not entirely by design, it is significant that the papers collected here constitute a representative survey of contemporary concerns and debates surrounding cognitive science. We believe that the volume will contribute to these discussions while also stimulating

further interest in these exciting developments in Australia and New Zealand, wherever this antipodean region might ultimately be located in the logical cartography of the field.[2]

PETER SLEZAK and W. R. ALBURY

NOTES

* The editors are very grateful to Anne Warburton and Anita Soekarno for their assistance in the preparation of this volume. The support of Apple Computer Australia is also gratefully acknowledged.

[1] D. C. Dennett (1986) 'The Logical Geography of Computational Approaches: A View from the East Pole', *The Representation of Knowledge and Belief*, M. Brand, and R.M. Harnish, (eds.), University of Arizona Press.

[2] In addition to the present volume, a further set of papers arising from the 1985 Cognitive Science Symposium has been published in W.R. Albury and Peter Slezak (eds.) (1988) *Dimensions of Cognitive Science*, Centre for Cognitive Science (University of New South Wales).

TABLE OF CONTENTS

FOREWORD vii

PREFACE ix

PETER SLEZAK / Introduction 1

J. A. FODOR / Why There *Still* Has To Be a Language of Thought 23

D. H. MELLOR / How Much of the Mind is a Computer? 47

KIM STERELNY / Computational Functional Psychology: Problems and Prospects 71

BERNARD BEROFSKY / Belief and Responsibility 95

CHRIS MORTENSEN / Mental Images: Should Cognitive Science Learn from Neurophysiology? 123

PETER SLEZAK / How *NOT* to Naturalize the Theory of Action 137

PHILIP CAM / Notes Toward a Faculty Theory of Cognitive Consciousness 167

MICHAEL A. ARBIB / Modularity, Schemas and Neurons: A Critique of Fodor 193

HUW PRICE / Action Explanation and the Nature of Mind 221

INDEX OF NAMES 253

PETER SLEZAK

INTRODUCTION

While the papers in this volume were written independently of one another, there is a clear overall unity in their concerns, with a few common threads discernable throughout. These reflect issues of central interest within cognitive science and include in particular the computational view of the mind. The purpose of this introductory essay is to act as a guide to the contributions by summarising their central arguments and also offering some brief commentary where appropriate.

1. J. A. FODOR: 'WHY THERE *STILL* HAS TO BE A LANGUAGE OF THOUGHT'

As the title suggests, Jerry Fodor is returning to defend the thesis of his landmark book *The Language of Thought*.[1] In this book Fodor provided the invaluable philosophical service of articulating clearly the sense in which mental or cognitive processes are to be seen as computational. For all the loose talk of the 'computer metaphor' for the mind, Fodor spelled out in detail the reasons for taking this view seriously and literally – essentially because all our most plausible psychological theories are committed to construing cognitive processes as computational. That is, Fodor pointed out that, whatever their details, the general structure of theories of cognition presuppose underlying computational processes and, specifically, a representational system in which these processes are carried out. Theories of deciding, perceiving, learning, language acquisition, all presuppose that the organism has available a medium of internal representation of great richness and this representational system must have properties similar to natural languages. In this sense Fodor argued for the existence of a 'private language' though one not subject to Wittgensteinian concerns. For as Fodor noted,[2] whatever Wittgenstein is supposed to have shown to be impossible, Fodor's theory posits something quite different and more akin to the machine language of a computer. After all, he noted "there *are* such things as computers and whatever is actual is possible". Thus Fodor relied heavily on the machine analogy in a number of respects which were explicitly spelled out. He wrote:

Peter Slezak and W.R. Albury (eds.), Computers, Brains and Minds, 1–21

When we think of an organism as a computer, we attempt to assign formulae in the vocabulary of a psychological theory to physical states of the organism (e.g., to states of its nervous system). Ideally, the assignment should be carried through in such a fashion that (some, at least) of the sequences of states that are causally implicated in the production of behaviour can be interpreted as computations which have appropriate descriptions of behaviour as their 'last line'. The idea is that, in the case of organisms as in the case of real computers, if we get the right way of assigning formulae to the states it will be feasible to interpret the sequence of events that *causes* the output as a computational *derivation* of the output. In short, the organic events which we accept as implicated in the etiology of behaviour will turn out to have two theoretically relevant descriptions if things turn out right: a physical description by virtue of which they fall under causal laws and a psychological description by virtue of which they constitute steps in the computation from the stimulus to the response.[3]

In this attempt to capture the appropriate level of theory at which the computational states and relations can be expressed, of particular importance for Fodor is ordinary belief-desire talk or the propositional attitude talk of 'folk psychology' which will constitute by and large the correct level. That is, believing, desiring, fearing, hoping, intending, learning, perceiving, etc., will be explicated in terms of some corresponding computational relation to a formula of the internal code. Fodor has thus been an advocate of the indispensability of commonsense psychological explanation, specifically because of its implicit commitment to the intentional contents of mental states and their causal role in the aetiology of behaviour and other mental states. This is what Fodor refers to in his present paper as 'intentional realism' and it is important to note in passing that not all philosophers share this realist view of the propositional attitudes. Dennett, Stich and the Churchlands, for example, are skeptical about this matter for a variety of different reasons.

Fodor's claim that there is a language of thought, however, goes significantly beyond the mere commitment to intentional realism. That is, the language of thought (LOT) thesis goes beyond the claim that there are internally represented mental states which have a causal role in behaviour to the further claim that these internal states also have a complex constituent structure which is appropriate to their semantic content through mirroring the structure of their associated objects.

The clarification and defence of this thesis with its 'classical' computational conception of the mind has become particularly urgent in recent times with the dissent growing from the direction of 'new connectionism'. Fodor's view can been seen as favouring the Turing/von Neumann

computer architecture as appropriate to explaining the structure of the mind, and the arguments here have recently been redeployed against the claims of 'connectionism'.[4] The claim of LOT for structured mental states is that beyond having a causal role and being semantically evaluable, the mental states must also have sub-parts which are themselves semantically evaluable. These sub-formulae will be the atomic components making up complex formulae and will provide the explanation for the commonalities among different molecular formulae which would appear accidental and utterly inexplicable on a non-compositional account. Fodor has called this property of mental representations their 'systematicity', which refers to the idea that the ability to produce or understand a sentence is intrinsically connected to the ability to produce or understand certain others – an ability which can only be accounted for on the assumption of a combinatorial or constituent structure for mental representations. In this respect, the systematicity of thought is exactly like the systematicity of language and explains the connection between our ability to think certain thoughts and our ability to think certain other related ones. The intentional realist's mere postulation of semantically evaluable internal states (perhaps neurological states) does not go far enough, since being in one such state need be in no way connected with being in another semantically related one. Fodor's argument here for a language of thought rests on the need to explain the remarkable contingent fact of the existence of connections among thoughts. In Fodor's view, the overall situation is quite simple: the need to explain such central phenomena means that "As things stand now, the cost of not having a Language of Thought is not having a theory of thinking".

2. D. H. MELLOR: 'HOW MUCH OF THE MIND IS A COMPUTER?'

It is on this last point that Mellor's paper provides a valuable contrast, for he argues precisely against the broad generality of a computational view of mind. Specifically, Mellor's thesis is that, while the limited domains of perception and inference which involve belief may be computational, the rest of the mind, and therefore most of it, is not. Details aside, as a general methodological matter, Mellor's view has the disadvantage of complicating the story of the mind through requiring radical discontinuities between the mechanisms subserving belief and everything else. *Ceteris paribus*, we would expect uniformity in the fundamental processes underlying cognition. In this respect Mellor's

view confronts a difficulty not faced by the more thoroughgoing rejection of Fodor's computationalism to be seen in Patricia Churchland.[5] Churchland rejects the entire 'sententialist' approach as a model of cognition, tied as it is to the concepts of language and folk psychology, whereas Mellor concedes its application at least for truth bearing states to be found in perception, inference and belief.

In agreeing with the need to postulate propositional attitudes, Mellor avows an intentional realism, but it is precisely in the degree to which he refrains from going further that he departs from Fodor and the LOT view. Fodor's computational account turns on construing the propositional attitudes as having (a) a semantically evaluable propositional content and (b) an attitude such as belief or desire which is taken to be a computational relation between the organism and the proposition. Mellor does not accept the generality of this view and accepts this analysis for only those propositional attitudes involving belief in some way. Mellor articulates an account here in which he emphasizes the syntactic basis of computation and its underlying causal processes turning on specific intrinsic properties of the tokens involved. This is what Fodor has called the 'formality condition'.[6] However, Mellor presses the question: 'When is a causal process a computation?' Mellor's answer is that the criterion for a causal function's being syntactic depends on semantics. In this sense, Mellor argues that a purely formal syntax is not autonomous, but presupposes a semantics and, in this sense, the semantics does not merely exploit syntactic properties but actually creates them. Thus, Mellor suggests that it is only when we know where the semantics comes from that we will know what it takes to be a computer and, in particular, how much of the mind may be one.

In the case of computers, Mellor points out that it is obvious the semantics derive from us, but it has been precisely the asymmetry in this regard which has been seen as the fundamental problem for understanding intentionality, content and how the mind might be seen as computational. In particular, the apparent discrepancies between folk-psychological type-individuation of mental states (so-called 'wide psychology') and the purely syntactic or formal individuation ('narrow psychology') has been a source of on-going perplexity and debate.[7]

For Mellor it is essential to distinguish those causal processes which are truly syntactic and computational through having a semantics, from those which are merely non-representational processes which might

nevertheless permit an algorithmic description. Thus, Mellor is relying on a distinction which is familiar from debates notorious in connection with grammars and their 'psychological reality'. In this context, too, a distinction has been pressed claiming that a system may be describable in terms of rules which nevertheless it may not be said actually to follow. Here Mellor argues that mental processes other than perhaps perceiving and inferring are not computational because the states involved do not represent anything in the required sense. Thus, sensations and pains, for example, do not represent their causes in the way that a belief might and are more like the case of a mass which merely functions in accordance with Newton's laws but cannot be said to use them to work out a resultant force. It is here that Mellor confronts Fodor's LOT claims most directly. He takes an example which happens to be a paradigm case for Fodor and, indeed, precisely one used in the original book[8] to argue the case *for* the inescapability of a computational account. Relying on the distinction just noted, Mellor suggests that decision theoretical accounts of action in terms of some calculation of expected utilities etc., are plainly false. For, he suggests, although one *might* conceivably act through performing such a decision-theoretic calculation, nevertheless one plainly does not. Mellor even concedes that such a model may accurately capture how beliefs and desires cause me to act, but he says "That doesn't mean I compute my action from them, and I don't". Here Mellor seems to be relying on some distinction between what I am aware of doing and what is unavailable to consciousness, but this hardly seems adequate to deny the claim that the underlying causal processes are computational, albeit unconscious. Reliance upon the comparison with a mass obeying Newton's laws seems particularly questionable in the present case, where we do, after all, have good reason to attribute the causes of action to internal representations. In this case it is not at all clear what could be meant by saying that we simply react in the way a mass does rather than act as a consequence of computational processes. It is worth noting that Mellor's position here relies on the kind of anti-realism adopted by the critics of Chomsky in relation to the status of the formalisms of competence theories in linguistics. In response, Chomsky's repeated question has been to ask why we should not take the implications of our best theories seriously and literally when they make certain posits, just as we ordinarily do elsewhere in science.[9] Of course, this recalls Fodor's essentially similar strategy in the original

argument for a computational view of LOT,[10] where he argued that we are provisionally committed to a computational view insofar as we take our best available theories seriously.

Aside from an instrumentalism regarding the status of formal models, Mellor's argument depends principally on his claim that, besides belief, the other propositional attitudes do not embody truth-bearing propositions. Hoping, fearing or wanting does not represent some fact as obtaining. Notice that Mellor is making a much stronger claim than merely adverting to the content-bearing character of propositional attitudes, for it is not just the semantic evaluability he points to as relevant. Mellor appears to require that the mental states be not only semantically evaluable, but assertive in the sense of actually purporting to be true. Commands and questions, for example, fail on this score since, Mellor claims, they don't embody a proposition representing some fact as obtaining. Mellor's argument here seems highly counter-intuitive, for the mere having of representational content and thereby truth conditions is not to assert their obtaining. Nevertheless, it is this failure which Mellor takes to disqualify the propositional attitudes other than belief from embodying information and being computational. Even if Mellor is right on this point, it may not impugn a computational account of mind *per se*, but only this particular view of what it is to be computational. There are, of course, quite independent grounds for construing the information processing of the mind as computational at a level more remote from the propositional attitudes of folk psychology and closer to the activities of the neurones as, for example, in the case of the retina or the regular architecture of the cerebellum, to cite only two well-known examples in which the computational processes are becoming reasonably well understood.

3. KIM STERELNY: 'COMPUTATIONAL FUNCTIONAL PSYCHOLOGY'

Sterelny too takes up the computationalist thesis, considering it together with the closely allied notion of homuncular functionalism. The latter advocated by Dennett as a kind of modularity is usefully contrasted with Fodor's recent, though quite different, modularity claims.[11] Sterelny introduces grounds for skepticism about the thoroughgoing functionalist idea that there is a purely abstract, autonomous level of description independent of neural or other realization. He suggests that some functional states, like sexual desire, may well be more intimately

coupled with their biological realizations. The advantage of homuncular functionalism over previous unitary accounts is in its providing for multiple levels of psychological description.

Sterelny points out that the idea of cognition as computation has been closely connected with functionalism. Indeed, perhaps more closely than he suggests for, in fact, the earliest version of functionalism due to Putnam was stated in terms of Turing machines and in this form gave a precise sense to the idea of the abstractly specified states of a device, independent of its realizations. Sterelny enumerates many of the virtues of the computational point of view, noting in particular a point we have seen raised by Fodor and Mellor, namely, the way in which causally related states subserve the information processing transformations. Sterelny points to the fact that the information processing functions performed by our brains have no access to distal causes, but only to intrinsic properties such as the neuronal signalling patterns. This is the problem which has continued to cause great consternation among philosophers, since intentional states appear to require going beyond purely formal or syntactic properties and adverting to semantic properties such as truth and reference. Sterelny's way of putting the point here in terms of the brain's exclusive access to intrinistic, formal properties is a useful antidote to prevailing skepticism about the adequacy of a purely syntactic or 'narrow' psychology to be seen in Curge[12] and others. Certainly when the formal processes of a computer are considered, their interpretation or meaning derived from our own design and purposes leaves obscure how mental states in us could have intrinsic content. Nevertheless, it would seem that whatever the shortcomings of the computational view in other respects, this 'formality condition' or 'methodological solipsism' ought not to be among them, since neuroscience must be just as formal and solipsistic in this regard. However we are to make sense of the semantic, intentional states of minds, here at least there is a symmetry with computers.

The motivation for a 'wide' psychology to accomodate semantic properties of intentional states comes largely from folk psychology. In the course of considering the dispute about the role of folk psychology in cognitive science, Sterelny notes that folk psychology is semantic psychology in the sense of positing states with contents about the world. As just observed, it is this which leads to the apparent conflict between folk psychology and the computational point of view. However, Sterelny point out that whatever the ultimate fate of folk psychology, it

need not threaten homuncular functionalism which posits a hierarchy of levels in the analysis of mind. As far as the computational point of view is concerned, Sterelny draws the moral of the conflicting perspectives in asking what would be left of the view that cognition is symbol processing if semantic properties do not find a place in a mature cognitive science.

Sterelny goes on to raise the question of computationalism in the form of a concern about its biological plausibility. Among the central issues here is the claim by Fodor originally in his *Language of Thought*[13] concerning the process of concept acquisition. On the face of it, Fodor's claim is startling and implausible. Baldly stated, it is the view that we can only acquire those concepts already represented in our innate language of thought and in this fundamental sense we can never really learn new concepts. Patricia Churchland[14] has drawn out the extremely counter-intuitive consequences of Fodor's view, and in taking up these criticisms here, Sterelny considers possible avenues of salvation for Fodor's thesis via the notion of 'triggering' which can render innateness claims more subtle and correspondingly more plausible. A further basis for Churchland's skepticism about Fodor's claim for a language of thought is what she dubs the 'infralinguistic catastrophe'. This is a reference to the fact that there are intimate continuities between human language users and non-linguistic animals or pre-linguistic children. This suggests the implausibility of any account of cognition which is too closely modelled on the properties of language. It is suggested that this is particularly so in view of the late arrival of language in the evolutionary process. In discussing this problem, Sterelny makes the elegant point that this argument can be turned on its head, for it is likely that language should have inherited fundamental properties of the systems from which it must have evolved. Nevertheless, overall Sterelny is sympathetic to the biologically based criticisms of computationalism and concludes that while this may well be wrong, the functional theory is more certain to be vindicated.

4. BERNARD BEROFSKY: 'BELIEF AND RESPONSIBILITY'.

Bernard Berofsky takes up central problems we have seen raised in the preceding discussions concerning the relation of folk psychology to a prospective cognitive science. In particular, he considers the paradigmatic propositional attitude *belief*, which has a particular interest beyond its centrality in the foregoing regard – namely, its essential role in