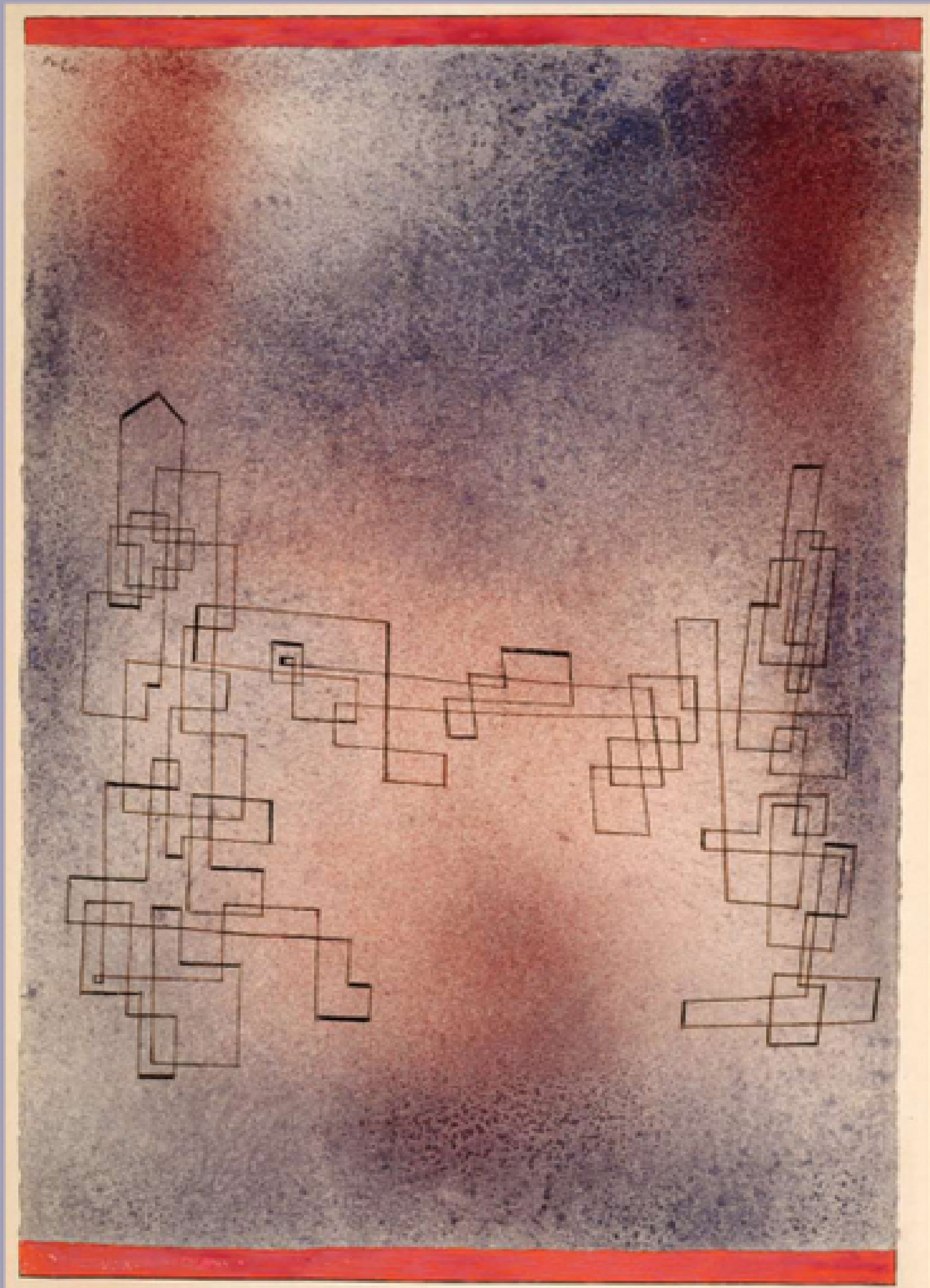
HISTORY OF COGNITIVE NEUROSCIENCE



M. R. BENNETT AND P. M. S. HACKER

WILEY-BLACKWELL

Table of Contents

History of Cognitive Neuroscience

M. R. Bennett and P. M. S. Hacker

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

This paperback edition first published 2013

Edition history: Blackwell Publishing Ltd (hardback, 2008)

Blackwell Publishing was acquired by John Wiley & Sons in February 2007. Blackwell's publishing program has been merged with Wiley's global Scientific, Technical, and Medical business to form Wiley-Blackwell.

Registered Office

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

Editorial Offices

350 Main Street, Malden, MA 02148-5020, USA

9600 Garsington Road, Oxford, OX4 2DQ, UK

The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, for customer services, and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

Library of Congress Cataloging-in-Publication Data

Bennett, M. R.

History of cognitive neuroscience / M.R. Bennett and P.M.S. Hacker.

p. ; cm.

Includes bibliographical references and index.

ISBN 978-1-4051-8182-2 (hardback : alk. paper) ISBN 978-1-1183-4634-1 (pbk. : alk. paper) 1. Cognitive neuroscience–History. I. Hacker, P. M. S. (Peter Michael Stephan) II. Title.

[DNLM: 1. Cognitive Science–history. 2. Neuropsychology–history. 3. Brain–physiology. 4. Cognition–physiology. WL 11.1 B472h 2008]

QP360.5.B463 2008

612.8’233–dc22

2008018526

A catalogue record for this book is available from the British Library.

Cover image: Paul Klee, Threatening Snowstorm, 1927, 291, pen and watercolour on paper on cardboard, 49.9 × 31.6 cm. Scottish National Gallery of Modern Art, Edinburgh, Vermächtnis Anna G. Blair. © DACS 2008.

Cover design by Nicki Averill Design.

For Adam and David

List of Figures

Fig. 6.12 Identification of the motor and somatosensory cortex in primates and humans due to Sherrington, Penfield, and their colleagues

List of Plates

Foreword

Anthony Kenny

Max Bennett and Peter Hacker's *Philosophical Foundations of Neuroscience* (Blackwell, 2003) was the most significant contribution to philosophy of mind in recent years. It examined thoroughly and carefully the pretensions of cognitive science to have superannuated philosophical psychology. It showed how the writings of some of the most prominent proponents of the new discipline are infected throughout with philosophical confusion. Those who scorn our ordinary concepts of thought, intention, and reasons as relics of a folk psychology, are engaged, as Bennett and Hacker showed, in sawing off the branches on which any scientist exploring the neurological basis of the mind must have to sit.

Since many a substantial reputation had been built upon the pretensions thus questioned, it was unsurprising that the book encountered hostile reaction, in particular from philosophers who had invested in cognitive science. Two of the best known, Daniel Dennett and John Searle, elaborated severe criticisms at a meeting of the American Philosophical Association in New York in 2005. The objections and responses from this session were published in 2007 by Columbia University Press as *Neuroscience and Philosophy,* a volume that enabled the world at large to see that Bennett and Hacker had more than held their own against their critics. But there remained in the minds of some readers a lingering suspicion that the two authors were not as close as they should have been to the technical work in cognitive neuroscience.

The present volume should set these doubts definitively at rest. Against a broad historical background it sets out in detail the research papers since Helmholtz in the nineteenth century that are considered to have established the discipline of cognitive neuroscience. In each section the authors first describe the relevant investigations, next subject to critical analysis the interpretations that the investigators placed on their findings, and finally restate the conclusions of the research in terms that have been purged of philosophical confusion.

Thus, to take one example, the authors describe the functional disorders consequent on commissurotomy, which led Gazzaniga and Le Doux to claim that each of the two hemispheres of the brain has a mind of its own. They object to this claim as an instance of the mereological fallacy: only a complete human being, not any part of one, can have a mind. They then restate the research results in neutral terms: the transmission of neural signals across the corpus callosum is a necessary condition for a person to know what is visually presented to him.

Another example: the examination of the memory abilities of patients with damaged temporal lobes led a number of investigators to conclude that memory involved the storage, in the hippocampus, of neural representations of past experiences. Bennett and Hacker show that this involves a misuse of the concept of *representation.* What the studies have shown, they claim, is simply that but for certain neural configurations or strengths of synaptic connections one would not be able, for example, to remember the date of the Battle of Hastings.

The reader's first reaction to Bennett and Hacker's restatement of research results like these may be that they take all the glamour out of them, presenting them in black and white, as it were, instead of in technicolour. But, in fact, these restatements are much more deflationary than that:

they not only concern the presentation of the story of cognitive neuroscience but have a significant effect upon the plot itself.

According to Bennett and Hacker, the besetting sin of neurophysiological researchers has been the attribution to brain parts and brain processes of states and activities that are logically ascribable only to entire animals. It makes no sense, they insist, to ascribe to parts of a creature such psychological attributes as being conscious, thinking, believing, perceiving, hypothesising, knowing, or remembering.

Such attribution defeats the explanatory project in two different ways. On the one hand, it offers illusions of explanation when no explanation has been given. The idea that there is a stored representation available to a person that makes him able to remember presupposes memory and cannot explain it: for were such records available to us we would still have to remember how to read them. On the other hand, the mereological fallacy also throws up questions that are only pseudo-problems, such as how information carried by different neural pathways enables an animal to perceive a unified object (‘the binding problem’). The critical stance of the present book presents no threat to neuroscientific research: it only averts futile questions that can have no answer, and deflates hype that goes beyond the empirical results.

In the background of this history of the relationships between philosophy and physiology two intellectual giants stand out: Aristotle and Descartes. In physiology Aristotle’s influence was malign and Descartes’ was benign; in philosophy the situation is reversed. Many of the functions of the brain were erroneously attributed by Aristotle to the heart; fortunately it was not long before the brain was given its rightful place by Galen. Descartes, however, made substantial contributions to neurophysiology and, if we are to believe Bennett, his insistence that biological explanation must be in terms of efficient causation was the foundation of all the advances in neurophysiology since the seventeenth century. On the other hand, Descartes' philosophical dualism threw philosophy of mind into utter darkness, and his shadow is so long that contemporary materialists still believe in a Cartesian ego, merely identifying it with the brain rather than with the mind. Bennett and Hacker constantly recall us to the Aristotelian concept of the unitary human being, recently given magisterial restatement by Wittgenstein.

In this book the findings of neurophysiological research are presented in an original and unfamiliar light. This is not because of fresh empirical investigation or new experimental work: it is the result of logico-linguistic analysis of the concepts involved in exploring the relationship between the mind and the brain. This style of inquiry will be unfamiliar to many physiologists, but it is a tool that is essential to the research project. 'The moral of our tale,' Bennett and Hacker say, 'is that neuroscientists need to devote as much care to ensure conceptual coherence and lucidity as they do to the experiments they undertake.'

Acknowledgements

Portions of chapters in this book have been published by us in the following journals and books:

Chapters 1 and 3: 'Perception and memory in neuroscience: a conceptual analysis'. *Progress in Neurobiology* 65 (2001), 499–543. Copyright (2001). Reprinted with permission of Elsevier Science.

Chapter 4: 'Language and cortical function: conceptual developments'. *Progress in Neurobiology* 80 (2006), 20–52. Copyright (2006). Reprinted with permission of Elsevier Science.

Chapter 5: 'Emotion and cortical-subcortical functions: conceptual developments'. *Progress in Neurobiology* 75 (2005), 29–52. Copyright (2005). Reprinted with permission of Elsevier Science.

Chapter 6: 'The motor system in neuroscience: a history and analysis of conceptual development'. *Progress in Neurobiology* 67 (2002), 1–52. Copyright (2002). Reprinted with permission of Elsevier Science.

Chapter 7: 'The conceptual presuppositions of cognitive neuroscience: a reply to critics', in *Neuroscience and Philosophy: Brain, Mind and Language* (2007) by Maxwell Bennett, Daniel Dennett, Peter Hacker and John Searle, pp. 127–62. Copyright (2007). Reprinted with permission of Columbia University Press.

Introduction

Neuroscience is concerned with understanding the workings of the nervous system, thereby helping in the design of strategies to relieve humanity of the dreadful burden of such diseases as dementia and schizophrenia. Cognitive neuroscientists, fulfilling this task, also illuminate those mechanisms in the brain that must function normally in order for us to be able to exercise our psychological faculties, such as perception and memory. Cellular and molecular neuroscientists study such mechanisms as those involved in the propagation of the action potential in neurons and their axons and investigate how this potential change releases transmitter substances from the axon terminal onto closely apposing neurons at the synapse. (For the development of this subject see Bennett 2001.) Cognitive neuroscientists are especially interested in how networks of synapses operate to fulfil their functions in the brain. Such networks, consisting of thousands to millions of neurons, each possessing up to 10,000 synapses, are found in parts of the brain that must function normally in order for us to be able, for example, to, remember a novel event for more than about one minute (the hippocampus) and to see (the retina and primary visual cortex V1).

For the last two decades, the mechanism of operation of the hippocampus has been a major focus of neuroscientific research. The neuron types to be found in the hippocampus, their spatial distribution and synaptic connections were first described by Ramon y Cajal (1904). A favourite approach to understanding hippocampal functioning is to develop what is taken to be a neural synaptic network representation, reflecting an engineering approach to the problem. Brindley (1967) suggested that some synapses in the hippocampus and therefore in its synaptic network representation should be considered modifiable. What he meant by this term is that the synapses are able permanently to change their properties following the arrival of an action potential in the axon terminal, so that in the hippocampus ‘conditioning and memory mechanisms of the nervous system store information by means of modifiable synapses’ (p. 361). Later, Marr (1971) suggested that ‘the most important characteristic of archicortex (hippocampus) is its ability to perform a simple kind of memorizing task’ (p. 23). He went on to show that the pattern of synaptic connections in certain parts of the hippocampus indicated that it could function as an autoassociative memory if the efficacy of the excitatory synapses were modifiable and if the membrane potentials of the large neurons were set by inhibitory interneurons that measure the total activity of the synaptic network. Marr’s suggestions, framed in engineering terms, were very influential. For example, in one account of how such neural network representations of the hippocampus work, it was suggested that ‘the recall of a memory begins with the firing of a set of pyramidal neurons that overlap with the memory to be recalled’ and that ‘the firing of different sets of pyramidal neurons then evolves by discrete synchronous steps’ until the stored memory pattern of neurons is retrieved (Bennett et al., 1994, p. 167).

The reduction of parts of the brain to engineering devices such as neural networks in the past half century or so has been accompanied by a major movement in the cognitive neurosciences: namely, that of taking the psychological attributes that are normally ascribed to humans (and in some cases to other animals) and attributing them to neural synaptic networks, either before or after they have been reduced to engineering devices of varying degrees of complexity and modifiability. In this book, we examine the claims that particular synaptic networks or clusters of synaptic networks in the brain can see (chapter 1), attend (chapter 2), remember (chapter 3), understand, think and translate thought into speech (chapter 4), and have emotions (chapter 5). Our approach has been to illuminate the historical development of these ideas and how they have been incorporated into the accepted jargon of mainstream cognitive neuroscience by studying the experiments whose interpretation gave rise to them. By conceptual analysis we hope to have shown what is awry with the interpretations of eminent neuroscientists such as J. Z. Young (1978), who suggests that 'We can regard all seeing as a continual search for the answers to questions posed by the brain. The signals from the retina constitute "messages" conveying these answers (p. 119).' Blakemore (1977) contends that the visual cortex in the occipital pole possesses neurons that 'present arguments on which the brain constructs its hypotheses of perception' (p. 91). Zeki (1999, p. 2056) argues that 'the interpretation that the brain gives to the physical property of objects (their reflectance), an interpretation that allows it to acquire knowledge rapidly about the property of reflectance' is required for us to see colours. Gazzaniga and his colleagues (2002) postulate that 'The right hemisphere is capable of understanding language but not syntax' and that 'The capacity of the right hemisphere to make inferences is extremely limited' (p. 414). Critical scrutiny of the idea that synaptic networks in the brain possess psychological attributes is at the heart of the present work.

In our previous co-authored work, *Philosophical Foundations of Neuroscience* (*PFN*), we identified conceptual problems in important current neuroscientific theories concerning, for example, perception, memory and emotion. Contemporary writing on the nature of consciousness by both neuroscientists and philosophers also received critical appraisal. The present work is distinguished from *PFN* in that we here present a study of experiments that have given rise to the various claims over the past century or so concerning the relationship between brain function and our psychological attributes (hence the large number of illustrations). In addition, the present work provides the opportunity for us to respond to philosophical critics of *PFN* such as Paul Churchland, Daniel Dennett and John Searle, especially on the subject of the brain and consciousness (chapter 7).

We hope that the present work, being much closer to the experimental activity of cognitive neuroscientists than was *PFN*, will engage their interest and critical response. In this way a dialogue might be generated amongst the neuroscientists in addition to philosophers concerning what cognitive neuroscience might hope to reveal about brain function in relation to the exercise of our psychological faculties. It is only through an ongoing critical, analytical appraisal of the experimental observations and their interpretation by both philosophers and cognitive neuroscientists that the subject, stripped of hyperbole, will truly prosper.

1

Perceptions, Sensations and Cortical Function: Helmholtz to Singer

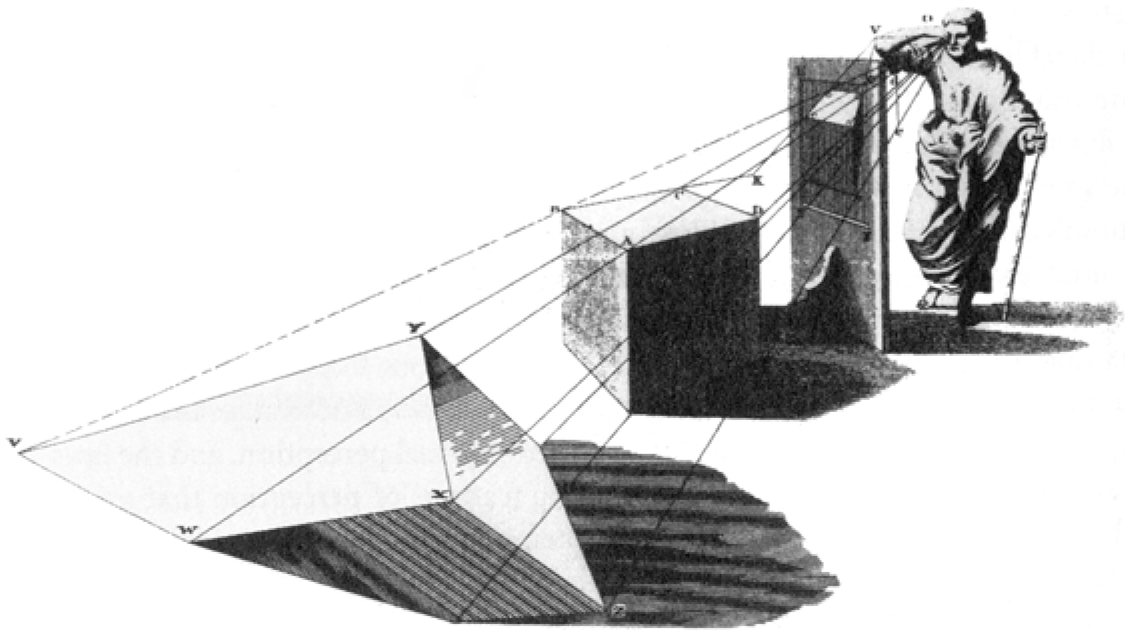
1.1 Visual Illusions and their Interpretation by Cognitive Scientists

Helmholtz (fig. 1.1), in his *Treatise on Physiological Optics*, suggested that the formation of a perception involves the development of an unconscious hypothesis based on inductive inferences gained from sensations. For him perceptions are conclusions of unconscious inferences the premisses of which are unconscious and (more or less) indescribable sensations and (unconscious) generalizations about the correlation between past sensations and objects perceived. The viewer shown in fig. 1.2 takes the strangely shaped object in the foreground, looked at with one eye, to be a cube because it has all the identifiable features along the line of sight that a cube has. On Helmholtz’s hypothesis inductive inferences are made by the person in fig. 1.2 on the basis of the sensations due to the rays of light from the object, and these support the most likely hypothesis: namely, the perception of a cube.

Fig. 1.1. Helmholtz. Sketch by Franz von Lenbach (1894).

Courtesy of the Siemens-Forum, Munich.

Fig. 1.2. Drawing to illustrate Helmholtz’s argument on how a perception is formed. (Glynn, 1999, p. 197.)



A variety of illusions (e.g. The Ponzo illusion, Kanizsa's illusion, the Ames Room illusion) have been taken as explicable in terms of Helmholtz's theory (Glynn, 1999). That is, these illusions can be explained by reference to the brain's drawing inferences from its past experience to form hypotheses about the objects of its present experience. In the Kanizsa illusion (fig. 1.3a) a ghostly white triangle emerges as a consequence of our inferring that this is the obvious way of interpreting the missing sectors in the three black discs and the edges of the black triangle. In the Ponzo illusion (fig. 1.3b) the upper horizontal bar looks longer than the lower one because the near vertical converging lines are interpreted as railway tracks with parallel lines receding into the distance. Another example of this alleged process of inductive inference is provided by the Adelbert Ames distorting room which produces the experience of extraordinary variations in size of people placed at different positions in the room (fig. 1.4). This room is constructed so that when it is viewed through an eyehole with one eye, an image is produced on the retina identical with that of a rectangular room of uniform height, whereas actually the far wall recedes and both the floor and the ceiling slope, as