

## PART IV

### Connectionism and the philosophy of memory

A good metaphor would be to imagine what the waves on the surface of a pond would look like after various objects had been dropped in the pond. Say the objects were an automobile tire, a beer bottle, the kitchen sink, and a telephone. The wave action on the surface of the pond shows traces of all but is specific to none. (B.B. Murdock 1982: 611)

It is characteristic of a creature, in contrast to a computer, that nothing is ever precisely repeated or reproduced; there is, rather, a continual revision and reorganization of perception and memory, so that no two experiences (or their neural bases) are ever precisely alike. (Oliver Sacks 1990: 49)

Preexisting knowledge, which often aids in the construction of elaborative encodings, can sometimes seep into and corrupt new memories. Such corrupting influences turn out to be a natural feature of many neural network models of memory. (Daniel Schacter 1996: 104)

## Introduction

In part IV I systematically apply to the case of memory some problems about mental representation which are much debated in theories of perception, and go on to defend the postulation of distributed memory traces against objections. Readers unworried by representations may be puzzled by the effort expended in evading dry philosophical complaints against ‘those “traces” that still plague psychology’ (Grene 1985: 43). But too many mistakenly take any invocation of representation to conflict with common-sense realism, with phenomenology, or even with neurobiology, and are thus tempted to dismiss out of hand the whole set of theories of memory I have described. These final chapters, then, are emphatically underlabouring toil: without clearing the ground by showing that distributed models (unlike static trace theories) do not suffer the conceptual incoherence with which critics charge them, no future and more positive interdisciplinary memory work will be secure. It is sad to have to be negative: but big questions about memory’s roles in the interweaving of self and world can barely be sighted while we are repeatedly told that a trace is not a memory.

Yet, despite the shift here to conceptual concerns, these chapters are also continuous with the historical material of the rest of the book. It is not so much that historical theories can solve modern worries about the veil of ideas between mind and world, as that historiographical differences in interpreting early modern concerns about representation reveal that apparently technical philosophical debates on traces in fact open up more general and urgent difficulties about the self and the past, or about truth in memory. This also indicates the role of culture: although modern philosophy of memory, obsessed by scepticism in epistemology, has yet had little interest in affect and culture, in asking why specific contexts of recall lead to particular reconstructions of the past, the new cognitive sciences can, if approached with an eye to society and history, transform the way problems are addressed in the philosophy of mind. This is all still promissory: at this stage I am removing obstacles to contextual cognitive science, rather than practising it.

So in chapter 15 I argue that distributed models allow the dissolution of traditional hostilities between ‘direct realist’ and ‘representationist’ approaches to memory, and that neither contemporary Gibsonian realism nor direct realist reinterpretations of the history of philosophy offer conclusive challenges to their promise. Chapter 16 centres on a new taxonomy of the many

criticisms of trace theories, and suggests that classical objections apply only, if at all, to local and not to distributed models. Finally, in chapter 17 I show how the motivating rhetoric of order and confusion in memory theory which permeates the historical critiques also applies in the methodological twists which connectionism provokes in cognitive science: memory is not a domain in which systematicity or regularity is always apparent.

Notoriously, Wittgenstein demands that we question the assumption that successful remembering must somehow correspond structurally to physiological regularities: 'Why must a trace have been left behind?' The expression of thoughts may be ordered: but 'why should this order not proceed, so to speak, out of chaos?' (1980: paras. 905, 903). From a quite different direction, Christine Skarda, a neurobiologist whose research centres precisely on chaotic brain processes, complains at Patricia Churchland that 'Brains do not represent. Representations are replaced by self-organizing neural processes that achieve a certain end-state of interaction between the organism and its environment in a flexible and adaptive manner' (1986: 187). Retention of the terms 'representation' and 'trace' is unimportant, save for polemic. But I want to show that their use in distributed models does not violate these strictures: although memory is not a notebook and 'the brain is not a writing' (Wittgenstein 1974: para. 131, 1982: para. 806), and although 'the word trace itself seems somewhat disreputable' (Goldmeier 1982: ix), it is still worth retaining it. Sandra Colville-Stewart's massive survey of physico-chemical analogies for memory concludes with the thought (1975: 402) that: 'it is the survival of the memory "trace" concept, some static, permanent, distinct storage form that each experience leaves in the organism, that links together most remarkably the oldest and most modern models'. If this is what traces had to be, then all hail to the critics: who would want them? But traces do not have to be static, permanent, or distinct, or even to be left by 'each experience'. What are conclusive objections to localist versions of the trace do not apply to the superposed traces characteristic of distributed models.

Problems of the self are raised by the historical consideration of representations and realism when, contrasting static and dynamic memory models, we realise that there really is a problematic 'third thing' in the old caricature of ideas barring subject from world. Ryle, Rorty, Grene, and all, encourage us to jettison the nasty intervening representations, hoping thus to rediscover the easy sense of being at home in the world which allegedly disappeared with Cartesian scepticism. But integration with nature need not be so homely: it is possible to argue that the troublesome aspect of the tripartite self/idea/world scheme is the subject rather than the mediating trace. There is a real world, a real past: and there are traces, patterns unfolding through time in complex causal relations with the world and with each other. And that is it: there is no further subject behind and separate from the traces at some deeper inner

sanctum (Dennett 1991a), no 'stranger behind the engram', in Daniel Schacter's luscious phrase (1982).

So the developing interdisciplinary connectionistic project needs to find more fragmented and empirically plausible ways of thinking of the self, which do not impose unity and coherence in advance. Memory is a test case for understanding the multiple constraints of body and culture which leave us unable simply to alter our selves at will. Temperament, history, chemistry, belief, and expectation all drive the omissions, inventions, hints and colour of particular acts of remembering (Bartlett 1932: 308–14), and these are precisely the kind of individual differences which make some individuals more susceptible than others to suggestion and false memory. These causes and courses of mixed associations extend to the society in which reconstructions occur (Connerton 1989: 37): 'It is not because thoughts are similar that we can evoke them; it is rather because the same group is interested in those memories, and is able to evoke them, that they are assembled together in our minds.'

If we do not simply possess passive memories, but are in part continually formed by them in ongoing contexts of use, then the sciences of memory must inevitably range across mixed natural and social environments as well as brains. This is how a thoroughgoing materialism actually encourages research on contexts and cultures. Remembering occurs in an inconstant world of 'leaks, holes, escapes and unexpected resistances' (Canguilhem 1966/1989: 198) a world which impinges on aware human bodies. These bodies in turn respond by fallibly shaping both their surroundings and their views of the past. Yet this past so often still drives them, unawares.