17
Order, confusion, remembering

> Memory can be compared with a storehouse only so far as it fulfils the same purpose. Where it does not, we could not say whether the things stored up may not constantly change their nature and so could not be stored at all.
>
> (Ludwig Wittgenstein, 'Notes for the "Philosophical Lecture"', MS 166 (1935/6), in Stern 1991: 204)

> . . . our personal identity is not simply a molecular formula of continual knowledge and skills; it is a singular compound of fragmentary systems of knowledge, incomplete stocks of information and discontinuous paradigms, disjoint fantasy fields, personal repetition cycles and intermittent rituals.
>
> (Alphonso Lingis 1994a: 148)

17.1 The vulnerability of distributed memory

I responded to the 'four-pronged fork' in chapter 16 by suggesting that distributed models with unsupervised learning rules can extract objective statistical regularities out of input deriving from the world without external control. But apart from the difficulty of scaling up from toy network models competent in restricted domains to psychologically realistic networks (Churchland and Sejnowski 1992: 125–37), the vaunted context-sensitivity of distributed models might go too far.[1] Just because they pick up these regularities in input, they are liable to being misled by ambiguities or idiosyncrasies of particular input sets. As Andy Clark, who calls this the 'hostage problem', says (1993: 171), 'the price of this sensitivity is an apparent vulnerability to the whims of a potentially hostile environment'. Networks which develop in good Gibsonian fashion by attunement to environmental regularities require 'the continued presence of a friendly training environment in which appropriate data are presented in an appropriate order' (1993: 173).

Connectionist practice demonstrates the difficulty of extracting relatively ordered partitions in representational space from certain kinds of input patterns. Elman, developing models of the representation of grammatical structure in which memory is 'inextricably bound up with the rest of the processing mechanism' and in which polysemy is a continuum with meanings differing naturally according to context (1990: 208, 191), had to reorder the training

1 Thanks to the audiences at cognitive science seminars at the universities of Melbourne, Sydney, and New South Wales for helpful discussions of earlier drafts of this material.

input so that more complex inputs were presented only later. In earlier versions, 'the network was unable to learn the task when given the full range of complex data from the beginning of training. However, when the network was permitted to focus on the simpler data first, it was able to learn the task quickly and then move on successfully to more complex patterns' (Elman 1989: 11–12). Exposure to full complex input from the start meant that no useful partitions were made, and the system could not learn. The problem exemplified here is that, without the artificial control of input complexity and order which Elman had to impose, distributed representational space will be too chaotic, too buffeted by complex inputs when the system has not yet developed internal structure of sufficient richness, to learn successfully over time. If this is so, Fodor's systematicity objection (chapter 11 above) gains force, and classical cognitivists can retort that only language-like symbolic structure will allow flexible capacities to develop in the face of poor or ambiguous input (Clark 1993: 172).

In response to the worry, connectionists can, firstly, design more powerful and stable networks, using recurrent nets and extra context layers to rival symbolic models in hard domains (Pollack 1990); and secondly, with Clark (1993: 173–88), look to evolutionary shields between networks and world. But historical attention to perceived dichotomies between order and chaos in human remembering supports the additional, complementary strategy of treating the vulnerability of distributed models as a virtue, to encourage re-evaluations of the explananda in memory theory. Where critics from Glanvill to Fodor require and impose intrinsic order in memory and cognition, supporters of superposition merely want their theory to allow for the emergence of order. Indeed thinking about the notion of *sequence* in remembering encourages suspicion about the very idea, or ideal, of order. The phenomenology of continuous flow between memories seems to gel with dynamic systems operating in real time, in which processes unfold and the variables change 'smoothly and continuously': in contrast, time in serial classical models 'is not real time, it is mere order' (van Gelder and Port 1995: 19–20). Connectionism does have a 'largely unnoticed capacity to illuminate our phenomenology' (Lloyd 1996: 61).

So it may be possible to bolster auto-associative mechanisms by clever design or with extra innate structures against environmental vagary. But just as useful is suspicion about critics' focus on abstract cognitive competences, their treatment of irregularities in performance as deviations from norms set out in tacit internal rules. Bechtel and Abrahamsen (1991: 254) complain that those 'impressed with the abstract regularities in behavior' captured in linguistic competence theories just talk past those whose concern is with variations in actual performance. In distributed models, a competence theory, expressing the structure of a task analysis, only approximates performance. As Smolensky puts it in the context of connectionist harmony theory, the competence theory gives the functions relat-

ing a system's input to an ideal output, whereas the performance theory is a 'differential equation governing the system's moment-to-moment evolution' (1988: 20). Regularity emerges from ongoing processing: but because it is not *guaranteed*, the theory will better extend to cover performance irregularities.

These points make sense of the historical debates on distributed memory. Glanvill's complaint that the ongoing processing of active memory trace motions and patterns, when they are not kept in a distinct storage system, will 'raise a little *Chaos* of confusion, where Nature requires the exactest order' (*VOD*: 36) and Coleridge's fears of 'the phantasmal chaos of association' (BL VII: 218) (chapters 5 and 14 above) are just older examples of the rationalist ploy of *requiring* regularity. Hartley acknowledged that in remembering 'perfect exactness is not to be supposed or required' (*OM* 375, I.iii.4, prop. 90), claiming that the many actual mistakes made in human performance make his distributed model of memory more plausible (chapter 13). In similar fashion, connectionists, attending to a fuller range of normal and abnormal cognitive and behavioural phenomena, seek to understand the common microstructure of cognition, and only afterwards ask if 'some order in the confusion might be forthcoming' (Churchland and Sejnowski 1992: 85).

Connectionists have already changed the explananda in specific psychological debates, on prototypes and exemplars, or on the acquisition of past tenses by English speakers, where initial connectionist modelling spurred new developmental research on just when children tend to stop over-regularising endings on irregular verbs (Rumelhart and McClelland 1986; Bechtel and Abrahamsen 1991: 194–200; Clark 1993: 155–62). Not only might actual performance be simulated without a mechanism governed by the explicit rules of the task domain: but the interaction of model and data can reveal that there is less order in later usage than was once thought.

The methodological strategy, then, is not only to challenge the task analyses of logicist cognitive science (Chater and Oaksford 1990), but to specify positive prospects for connectionist psychological explanation in domains where our cognitive capacities are not, in fact, all that systematic. Clark (1990: 195, 208) drives home the gap between competence and processing which connectionism opens up, where competence theories can only be idealised fixes 'on certain stable states of the system' which do not reveal 'the *dynamics*, or actual processing strategies'. Instead of implementing a prior theory of the structure of a task domain, connectionists must work backwards to trace the shape of a cognitive space *after* performance. Only thus do we 'avoid imposing the form of our conscious, sentential thought on our models of unconscious processing' (Clark 1990: 215–18).

Decentring requirements of order forces accounts of how cognitive regularity ever exists. 'How can we do logic if [our] basic operations are not logical at all?' (Rumelhart *et al.* 1986: 44): how does memory sometimes track past events

reliably? Some of the work is done by *external* structures and symbol-systems. When your theory suggests that there is chaos or mush inside, you have to look for order outside the mind/brain. Despite the tendency of Wittgenstein and his followers to reject all sciences of mind out of hand, this strategy is precisely in line with his cautions against projecting expressed regularities on to mysteriously structured inner processes. The 'system of impulses going out from my brain' when I talk or write need not itself 'continue further in the direction of the centre' (Wittgenstein 1980: para. 903; compare Davies 1991: 230–1; Shotter 1991: 205). Evolution and experience teach us to manipulate the environment and to supplement memory, turning hard cognitive problems into tractable ones: as Clark has recently put it, 'advanced cognition depends crucially on our abilities to *dissipate* reasoning' by building 'designer environments' in which 'the brain in its bodily context' can interact with a complex world (Clark 1997: 180, 191; see also pp. 59–69).[2] External symbols and recording devices, obviously, are culturally and historically specific, and the many senses of the past in different cultures and periods develop partly through different social uses of material and cultural artefacts and practices (Burke 1989; Lowenthal 1989; Radley 1990; Cole 1993; Munn 1995). So, again, it is not just that distributed traces are compatible with the attention to context, environment, and society which many critics of static traces demand: it is that the broadening of cognitive science beyond the mind/brain alone is specifically required by them.

17.2 Self, memory, confusion

One of the most obvious phenomenological features of memory seems untouched by all this cognitive and social scientism. The fears about memory's access to the real past which early modern thinkers projected on to animal spirits (chapter 9 above), their worry that they might be tricked by sly fluids into 'remembering' falsely, were about the intrusion of confabulation into memory. But there is a different subjective distinction available, relying on the sense of imagination which does not mean 'failed memory': do I not just know immediately when I am remembering as opposed to merely fantasising, when (as Reid would say) belief is inexorably attached? I may be misled in some way, but there is still a great difference between personal memory of an event and merely thinking that it happened.

Yet one secure finding in recent psychology of memory is that our 'remember/know' judgements, in deciding for instance whether we experienced

2 'One element frequently left out of cognitive modeling is the element of culture, that is, shared patterns of acquired behavior characteristic of a species. But the cognitive capacities of animals directly affect the kinds of culture they produce, and in the case of humans, the opposite is also true: specific types of human culture have direct effects upon individual cognition' (Donald 1991: 9–10; compare Hutchins 1995: 353–74).

something personally or just know that it occurred, are surprisingly malleable (Schacter 1996: 22–6, 114–21). Failures in 'source monitoring' (Johnson and Raye 1981; Belli and Loftus 1996: 174–5; Roediger 1996: 87–9) lead us to infer a likely cause in personal memory for beliefs, images, or other mental states which in fact derive from other sources. On one account, judgement of whether we subjectively remember something or just know it is determined not just by the real past experience but also, in particular, by fluency heuristics which make us attribute an image, say, to personal memory if it plausibly and quickly integrates with background knowledge (Kelley and Jacoby 1990, 1996). The conscious experience of remembering is not, as critics of trace theory complain, somehow contained in the trace itself: it is a result of constructive attribution in which conscious states can, but need not, be causally involved (Kelley and Jacoby 1990: 53; Mandler 1989).

Such constructions in turn influence future processing, and we can influence our memories and thus our selves *within* the causal nexus, abetted by whatever benevolent checks the social and natural environment affords. But the nature of the subjectivity and the control involved may not match pre-theoretical intuition. Intimately subjective beliefs about the relations between our current mental states and the past are no more transparent, no less accessible to science, than are beliefs about the voluntary origin of action, where the lack of immediate introspective awareness of the mechanisms involved does not mean that there are none: it is only 'the cognitive impenetrability of one's own actions which makes them appear autonomous and uncaused' (Slezak 1986: 431). The relevant sciences, again, will refer to the social as well as to the brain, for we spin tales about our past selves in linguistic and other narratives which suit particular purposes, shoring up some fragile coherence against pandemonium in story and autobiographical thoughts.

Neither neurophilosophy nor lab psychology alone will drive vast conceptual change, or force the elimination of common-sense self-conceptions. You can only revise what you can replace, and whirling revisionary schemes require independent motivation before evidence from cognitive science is to convince. But fragmentation has its seductions too (Churchland 1983: 85):

> If our conception of control changes, then our conception of the self changes *pari passu*, and this possibility may initially produce a kind of dizziness, tainted perhaps with anxiety. The warm antiquity of self, as Wallace Stevens calls it, binds us profoundly... The self, in the event, may turn out to be not so much a unified coherent entity, but a disconnected collection of wants, needs, and whatnots, flying in loose formation.

Regularity, order, and coherence in memory and cognition are fragile, temporary achievements: no less real for that, but more rare, less common, than we think.

The dream of memory as re-presenting the past in its full presence, anyway, may always have been more a philosopher's phantom than part of folk psychology. Experience was never fully captured even at the time, and the things 'stored up . . . constantly change their nature' in the superpositional mix. The act of autobiographical remembering obliterates by reducing the past to only one current context: but it also exceeds, as the early modern theorists struggling to keep the past in place realised, by creating a past which was never present. As one activity pattern flickers across brain pores, others subside, alter irretrievably, slip: but the animal spirits gather new wishes too in playing down their haphazard paths. Again, if traces are distributed fragments rather than monolithic containers, memory theory is not immune to loss, time, melding, and productive confusion.

This chapter has described the methodological shift which both connectionism and history encourage, and drawn on further psychological accounts of ordinary confusion, in which rigid control over memories is not required. This final part of the book has attacked the cogency of blanket refusals to countenance memory traces. Without prudish conceptual legislation, broader memory theory can perhaps recover tinges of both the sadness and the lightness of talk and theorising about the nimble animal spirits. We are, inevitably, more mixed within ourselves than we know.