12

Associationism and neo-associationism

> . . . what is stored is not in any sense any sort of item at all. An association is not a link or a path or a bond between two items. It is an unrecognizable aggregate of the two items it relates. (B.B. Murdock 1982: 625)

Introduction

In any associationist theory of memory, surely, there must exist separate local memories to *be* associated, 'some number of basic units' (Crovitz 1990: 167) to be related in experience. No! The rigour of empiricist associationism, surely, was in tying transitions between ideas to current stimuli, and excluding all 'cognitive voluntary factors' (Crovitz 1990: 168). No!

Of course most associationists, from Hobbes to Skinner, did accept these obvious foundational principles. But a quite different history of association lurks in the same traditions, and the connectionist return to dynamics in cognitive theory is neither just a triumph of trendy technology nor a blundering surrender to blunt behaviourism.

Atomism's consequences were sad. Bergson eloquently lamented 'the capital error of associationism' (1908/1991: 134):

> it substitutes for this continuity of becoming, which is the living reality, a discontinuous multiplicity of elements, inert and juxtaposed . . . the principle of associationism requires that each psychical state should be a kind of atom, a simple element. Hence the necessity for sacrificing, in each of the phases we have distinguished, the unstable to the stable.

Passive minds mechanically shuffle isolated bits of information, responding automatically to immediate stimuli, forever barred from true invention. Associationist minds were the victims of a dull environment with which they could not actively interact. There seemed no room for interesting analysis of interactions between the social and the psychological, for there could be no active cognitive state attuned to particular retrieval conditions: the properties of the dormant trace and of the bare input seemed together wholly to determine remembering, leaving no place for more complex retrieval dynamics (Schacter 1982: 165–70, 1996: 56–71).

Not the most valiant associationist would appropriate its *whole* history. Philosophers sympathetic to connectionism have made a range of responses to the charge that it is too close to old associationism to be a viable model of cogni-

tion. Some initially denied the link, but by tending to assimilate associationism to anti-mentalist behaviourism (Bechtel 1985). There are historical puzzles about just how associationism disappeared so wholly into behaviourism: but Hume and Hartley have more in common with Rumelhart and the Churchlands than either does with Watson and Skinner. More recently, Bechtel and Abrahamsen (1991: 102; compare Smolensky 1991: 202) suggest that connectionism is 'not a *return* to associationism; it is not *mere* associationism; but its most obvious ancestor is indeed associationism'.[1] This is a more subtle attitude: acknowledging ancestral relations allows genealogical excavation of aspects of the descent. Bechtel and Abrahamsen list technical developments 'not even conceived of' by classical associationists. They note that hidden units, mathematical models of the dynamics of learning, back-propagation, simulated annealing, and other specific learning mechanisms are new. But they also suggest (1991: 102) that connectionism, 'returning to the original vision of the associationists [and] adopting their powerful idea that contiguities breed connections', has 'an unprecedented degree of sophistication' because distributed representation has been conceived of for the first time. Andy Clark too (1993: 233, n. 1) takes connectionism's non-linear computational functions, 'which compress and dilate a representational space', to mark it off from classical associationism.

But the history in parts I and II already reveals a longer historical background to dynamic distributed memory *within* old associationisms. Before pursuing Hartley's development in chapter 13, I confirm that atomistic, localist associationism is not the only kind.

12.1 Meta-features of associationism

An analysis of conceptual foundations of associationism must be sufficiently general to encompass both twentieth-century learning theory and classical empiricist epistemology as special cases without collapsing the general framework into either. Early cognitivist setting of 'formal limits' on associationist explanation (the ancestors of Fodor and Pylyshyn's (1988) systematicity arguments against connectionism) tended to conflate associationism with behaviourism by allowing only possible descriptions of behaviour to figure as elements between which associations can hold (Bever, Fodor, and Garrett 1968: 583).

But there is a broader standard picture in characterisations by historians as well as philosophers. Initially opposing epistemological nativism and faculty

1 Compare van Gelder 1992b: 189–91: connectionism is 'in a deep sense associationist', but is immune to old criticisms because it denies that mind and memory are passive recording devices for independent atoms of information, and allows that minds continually construct environments, rather than forever reflecting them.

psychology, associationists tried 'to reduce faculties to aggregates of elementary sensory units', the union of which 'was accounted for in terms of mechanical connection, or a chemical analogy of compounding or fusion' (Young 1970: 95–6). It is possible to extract four broad meta-features of associationism, which I will then start to whittle down (compare Warren 1921: ch. 9; Young 1973; Anderson and Bower 1979: ch. 2; Fodor 1983, 1985a):

1 *Reductionism, elementarism,* or *atomism*: psychological structures are constructed from a set of elements (ideas, sense-data, memory nodes, reflexes) out of which psychological structures are constructed. Complex elements are decomposable into simple elements.
2 *Sensationalism*: simple elements are sensations.
3 *Connectionism*: elements are associated together through experience, with a relation of association defined over the elements.
4 *Laws of association*: mechanistic principles, in virtue of which experience determines what gets associated, are used to explain the properties of complex associative configurations by reference to the properties of the underlying elements.[2]

These meta-features are subject to a range of philosophical objections. I want to correct the first two before showing in some detail what goes wrong when they are assumed to characterise *all* associationist views.

12.2 Atomism and sensationalism

Psychological atomism, the view that the basic elements in or of the mind are primitive, unstructured, independent entities (whether sensations, sense-data, or raw feels) is *not* an essential part of associationism, and indeed conflicts with associationist talk of the fusion of memories, with chemical metaphors for association, and, especially, with the ever-present neuro-philosophical or psychophysiological strand in associationism.

Talk of 'simple elements', whether ideas or reflexes, by associationists is vulnerable to the charge that ultimate primitives, entities which have no complexity or internal structure, are far from obviously coherent (McMullen 1989, drawing on Anderson 1927/1962). Rumelhart, for example, refers to 'schemata' which are 'elementary in the sense that they do not consist of a further breakdown in terms of subschemata' (in McMullen 1989: 5). These look like Hume's 'simple perceptions or impressions and ideas' which 'are

2 Different versions of associationism are distinguished by the combinations of such principles which they allow. Possible mechanisms include spatial and temporal contiguity, resemblance or similarity, frequency or repetition, causation, inseparability, vividness, intensity or degree of strength, attention and degree of readiness, duration, recency, and so on. The mathematical learning rules of new connectionism are further examples.

such as admit of no distinction nor separation', and which make up the complex impressions and ideas which can 'be distinguished into parts' (*Treatise* I.i.1: 2). Such an 'ultimate simple' indistinguishable into parts would, it seems, 'have to be something which has no properties, i.e. nothing at all' (McMullen 1989: 6).

There are two linked responses to this point. Firstly, associationists are not now working with totally unstructured primitives: it is just that their 'elements' are not structured *syntactically* as are the combinatorially structured mental representations of Fodor's language of thought. No units or elements in associationist theory are *ontologically* basic as the logical atomists' ontologically neutral sense-data were. What are called associationist 'atoms' are psychological or psychophysiological and supervene on a physical base, and thus are not really atoms at all.

But an even stronger response to the problem about the incoherence of atoms is possible. In one sense, no *things* are associated at all in distributed models. There's no initial stage at which bare impressions float round the mind before they are hooked up with others: traces are always already complex. Effective encoding itself is elaborative, with new information rummaging and arranging existing memories, and old traces shot through with later accretions (Schacter 1996: 44–56). With no part of the system ever inactive, it often makes no sense to ask what items in memory *were* before they were associated with others: in dynamics, 'the beginning point and the endpoint of cognitive processing are usually of only secondary interest, if indeed they matter at all' (van Gelder and Port 1995: 14).

In turn, the second meta-feature of associationism can be challenged. Classical associationists' 'elements' were not always directly tied to sensations. Rather, they could be in systems which mediate between sensory input and behaviour. This is obvious in models of memory: the piled effects of past experience on the present state of the system and thus on current and future behaviour must work by way of long-term internal changes. In new connectionism, and in Hartley's version of classical associationism, relations with the environment are complex, some parts of the system being only remotely connected to the periphery. Of course new connectionists provide much more detail about mediating systems, but only if 'associationism' is assimilated to behaviourism and forced to appeal *only* to direct environmental stimuli will new connectionism appear conceptually new. To the extent that sensationalism *was* adopted by classical associationists, it can be attributed not to the psychological theory itself but to the epistemological anti-nativism of the empiricists (compare Warren 1921: 14).

The distancing of associationism from sensationalism has the advantage that psychological associationism gives no a priori answers about innateness. We cannot decide in advance on the roles of evolution and of early, past, and

present environments in teaching the organism. It is an empirical matter to discover which if any sequences of associations may be invariant, given evolution. Certainly, many philosophers of new connectionism retain an anti-nativist slant, and the focus on learning as a natural activity of the organism avoids many pitfalls of extreme nativism: but others consider the issues still more or less open (Clark 1987, 1993: 182–8; O'Brien 1987; Ramsey and Stich 1990).

12.3 Hampshire, Coleridge, and the con/fusions of memory

Reconstruing the history of associationism, then, suggests the dispensability of both atomism and sensationalism. Only *localist* forms of associationism have the tinge of blindness which subjugation to the stimulus brings (compare Bechtel and Abrahamsen 1991: 102). The ongoing dynamics of internal activity, in contrast, give distributed memories considerable independence from the present environment. Now I juxtapose another recent critic with a further strand in Coleridge's attack to reveal clearly the different criticisms which, though considering quite distinct forms of associationism, share a commitment to an active self above and behind the memories themselves.

In *Innocence and Experience*, an extended specimen of 'the enterprise of a moralist', Stuart Hampshire argues that the habit of 'dwelling upon the past' for its own sake is one of 'the distinguishing features of humanity' (1989: 32, 114). Lingering on memories which 'are mine alone' is one way to 'preserve the continuity of my experience and . . . confer some unity and singularity on my life as a whole' (1989: 114). It is this 'appropriately continuous history', rather than the singularity of an individual, which confers true personal identity (1989: 115). Hampshire goes on to seek an explanation for the value we attach to the individuality afforded by memory's 'personal particularity'. In this context he describes memory in terms 'intended to run counter to the famous implications of the association of ideas'. But Hampshire takes localist storehouse models as the pervasive paradigm in the sciences of memory, and so complains that *all* theories of memory are incompatible with the reconstructive nature of human remembering. I concentrate on his specific remarks about associationism.

Hampshire complains that classical associationist theories of memory could not catch the similarities between human memory and a compost heap, in which 'all the organic elements, one after another as they are added, interpenetrate each other and help to form a mixture in which the original ingredients are scarcely distinguishable, each ingredient being at least modified, even transformed, by later ingredients' (1989: 121). Associationistic memories are unnaturally independent of each another, original ingredients still distinguishable, 'their identity and integrity unmodified by their neighbours . . . like beads on a string' or individual stones in a heap. Memory metaphors should 'convey the unmechanical and confused connections which intimately

link our memories': Hampshire would prefer 'Heraclitean and William Jamesian metaphors of rivers and streams, which represent our memories fusing with each other to form our consciousness of our own past experience' (1989: 121).[3] There are, he argues, principled barriers to an associationist account of *causal holism* in memory, of how many different memories come together in a present act of remembrance, or of fusions in which encoded items might lose their identity and proliferate in a rich mulch of memory.

Hampshire's historical case is that associationism was *opposed* to the metaphors of rivers and streams which suggest this fusion. But this claim can be neatly refuted by remembering distributed memory: the leading associationist David Hartley suggested exactly such a picture. From Hartley's distributed model, worries Coleridge,

> [it] results inevitably, that the will, the reason, the judgment, and the understanding, instead of being the determining causes of association, must needs be represented as its creatures, and among its mechanical effects. Conceive, for instance, a broad stream, winding its way through a mountainous country with an indefinite number of currents, varying and rushing into each other according as the gusts chance to blow from the opening of the mountains. The temporary union of several currents in one, so as to form the main current of the moment, would present an accurate image of Hartley's theory. (*BL* VI: 215)[4]

This is exactly the metaphor for memory Hampshire recommends: but whereas he denies it to the associationist, Coleridge is well aware that the winding, gusting course of interfering distributed traces is just what Hartley's associationism catches. Coleridge ascribes to Hartley the very image of shifting control in a distributed model, with storage and processing entwined, which Hampshire had said was impossible for an associationist.[5]

3 This is odd: although James thought he had given 'abundant reasons for treating the doctrine of simple ideas or psychic atoms as mythological' (1890b: 552), he clearly did not think associationism was thus ruled out (1890a: 253–79, 1890b: 550–604).

4 This shows that chemical-fusion metaphors for association pre-date James Mill. Coleridge himself distinguished imagination from fancy in parallel terms: as Willey puts it (1949/1973: 24–5), where fancy merely juxtaposes existing images into mechanical mixtures in which the ingredients remain as they were when apart, imagination mingles its elements 'like chemical compounds . . . in which the ingredients lose their separate identities'. But Hartleian fusion occurs *without* the involvement of the will: it does not therefore reduce multiplicity to unity as Coleridge wanted, but confuses. See chapter 14 below.

5 Unintentionally, Hampshire's desiderata for a theory of memory actually read like a manifesto for new connectionism. My 'picture of the past' is 'confused, overlaid by accretions', my memories tend to interpenetrate and display a 'comparative lack of discreteness'. Memory is an 'apparently holistic device installed in the human brain'. He even uses 'network' and 'complex networking' to describe the system by which experience has its effects, and uses as an example 'the phenomenon of "take-off" in language learning' (a child's leap to sudden mastery in understanding and use of language) (Hampshire 1989: 122–3), on which some of the most hotly debated connectionist research has been performed. Hampshire's problems spring from his acceptance of the common assumption that holism is incompatible with mechanism.

It is just because of the contextual dependence of ongoing processing, as various causes conspire in temporary union to form 'the main current of the moment', that Coleridge in contrast sees 'the phantasmal chaos of association' (BL VII: 218) as *too* prone to fusion, too likely to destroy the identity of the original ingredients, too *con*fusing to be left undisciplined by will and reason. In rejecting associationism, Coleridge requires a central executive or cognitive control system to determine actively the ongoing processing of passive items of memory which are kept cleanly independent of will, reason, and judgement. Perhaps what is more surprising is that although Hampshire, like Fodor, characterises associative memories as too passive, where Coleridge saw them as dangerously active, he shares, in more moderate form, Coleridge's belief in or desire for a strongly role cognitive executive.

Hampshire's critique of some metaphorical and philosophical models of mind and soul is convincing.[6] But in rejecting distinctions which would privilege reflective thought over 'lower domains of thought', he retains a distinction between 'active and passive thought' (1989: 32–41). This distinction, which admits of degrees, is 'an incontestable phenomenon, and is not an invention of philosophical theory' despite the difficulty of formulating or analysing it (1989: 39).[7] Our activities or exercises of skill become thoughtless when 'we' have become passive, when 'one's thought strays or is distracted, and one is no longer directing thought to the task in hand'. It is in this context that Hampshire again refers to associationism, in similar vein. When 'we' are truly thinking, '[when we] assert ourselves and . . . exercise our power of thought . . . we seem not to allow our thought to drift onwards as the immediate effect of external causes and of the association of ideas' (1989: 40). Mere association, mere mental causation, then, makes us passive creatures of the environment and of ideas: the fusion of such ideas, when 'we' are not controlling its pleasant liquid flux, becomes confusion,

6 Trying to undermine pervasive philosophical 'dead metaphors' of mind, like that of 'the perpetual conflict of desires, the chaos among the unruly populace, which, like the mob in ancient Rome or in Renaissance Florence, needs to be mastered and controlled' (1989: 35), Hampshire argues that 'the mind, unlike the brain, does not literally have identifiable parts'. This means that 'all pictures or models of the human mind and of its faculties . . . are inventions for a philosophical and moral purpose, and they are all in this sense arbitrary' (1989: 35–6). It looks as if there are then no facts about the mind to be found. But after effectively challenging traditional views of 'the normative implications of the ideology of higher and lower faculties, and of the virtues connected with the obedience of the lower', Hampshire goes on to say that 'one can go behind the ideology and look at the facts'. These facts (about active thought, reflection, and control) have nothing to do with science or theory, but can only be salvaged 'in this context of moral speculation' (1989: 38). The urge towards executive control has many channels.

7 Hampshire's distinction is focused on activity of thought in practical rather than theoretical reasoning, refusing to privilege a concern with truth over a concern with practical skills or social/political engagements.

the surrender of 'the natural dominance of thought within the soul or mind'.[8]

The existence of 'active and directed thought' requires something to direct it, something perhaps identifiable with the first-person pronoun which Hampshire regularly uses. This is 'the thoughtful self' or 'the supervisory self' which looks down on what 'we' are up to, in 'phenomena of consciousness' which give 'a clear sense to the overlordship of thought in the soul' (1989: 40). Critics' opinions on how to treat the associationist have changed in two hundred years. Where Coleridge advised the use of 'discipline, not argument' on those despicable foreigners who denied the sovereignty of will (BL VII: 221), what is now recommended even by a philosopher who retreats from the primacy of reason, is the mild, if definite, discipline of 'stepping back' to supervise the present cognitive processes which 'we' own (Hampshire 1989: 40–1). The maintenance of sufficient order, of creative fusion instead of lawless confusion, requires for all of these critics the separation of the true executive which controls and supervises from the stored items which it manipulates.

Hampshire argues that the true pleasures and value of individuality lie in the contemplation, by this inner executive, of my 'spiritual capital', where that capital has the form of memories which 'are mine alone, the stuff of my inner life', memories selected out of the 'multitude of memories which we know to be ours and no one else's' (1989: 120, 114–15). 'We', unlike 'intelligent but non-human animals' which cannot progress past learned associations, both can and do 'cause [our] minds to linger among memories, or to cherish and to preserve their memories merely as memories' (1989: 114). Hampshire does not explain the nature of this strange causal interaction between an 'I' and a mind, an interaction said to occur whenever I pleasure myself in making it hover among and cherish its (my? our?) truly private memories.

It is too easy to carp, when there is still so much to do, on the oddity of a-empirical speculations. It is still hard to see that models of distributed memories as changes in interrelations between parts can explain how order does, sometimes, erratically, emerge at the psychological level from the phantasmal chaos of multiple nested con/fused memory traces. An actual theoretical model is needed: on, then, to Hartley.

8 Phrases like this suggest that Hampshire's critical awareness of the dangers of hierarchical models and metaphors has not been driven far enough. He is happy to acknowledge the normative implication of the distinction between active and passive thought: 'a person actively directing his or her thought is an autonomous agent, fully responsible for what he or she achieves, and in this respect to be praised' (1989: 41). This is an admirably clear demonstration of the intimate connection between issues about, on the one hand, cognitive order or chaos, and on the other, normative conceptions of agency, free will and free action, autonomy, and responsibility, which form the basis for moral praise and blame.