3
Memory and 'the Cartesian philosophy of the brain'

> In the true art of memory . . . by the leading back of things to causes, when all those things are finally led back to a single one, there will be no need of memory for any science. (René Descartes, *Cogitationes Privatae* (1619–21), AT x.230)

> There is a remarkable Story to confirm the Cartesian philosophy of the Brain. A Man hurt by the fall of a Horse forgot about twenty Years of his Life, and remember'd what went before in a much more lively manner than usual.
> (David Hume's early memoranda, Hume 1948: 502)

> It is a big mistake to accept the principle that no body moves of itself. For it is certain that a body, once it has begun to move, has in itself for that reason alone the power to continue to move. . .
> (Descartes to Mersenne, 28 October 1640, AT iii.213, CSM-K 155)

> Our machines are disturbingly lively, and we ourselves frighteningly inert.
> (Donna Haraway 1991a: 152)

3.1 Introduction

Descartes' psychophysiology is notoriously absurd, 'a little fantastical' (A 479), 'a baroque ruse' (Metraux 1996: 67) betraying exuberant rationalist disdain for observation. His neuromechanical speculations on the dramatic whirl of invisible spirits in hollow nerves are at best 'quaint', separable from his true metaphysical and epistemological achievements (Wilson 1978: viii; Schouls 1989: 154), at worst 'particularly uncontrolled' (Mackenzie 1992: 136; Hatfield 1992: 347–8). Reducing the phenomenological complexity of lived bodily experience to the atomic combinations of mythical particles, Descartes bypasses 'the concrete life of feeling' which he 'had done his best to avoid' (Grene 1985: 52). Since he barred the thinking subject from its body, which was just another object in a world 'not of meaning and love and laughter and tears . . . but of material particles going about their lonely business', it is no surprise that 'twentieth-century approaches to subjectivity are dominated by the anxiety not to be Descartes' (Rée 1995: 205–6). Specific microreductions of emotions and temperaments to diverse motions of animal spirits are 'intuitive but extraordinarily simplistic' (Gaukroger 1995: 273), and the physiological treatment of memory is 'particularly incoherent' (Richards 1992: 41). Descartes revels in wild hypotheses about body processes, entranced by the *ébranlement* ('commotion, disturbance') of nervous fluids through the folded mesh of the brain.

It is true that modern physiologist-historians, neuroscientists, and memory

theorists have repeatedly looked to Descartes for inspiration. The father of modern philosophy was already there (wherever present theory stands) in neurophysiology too! With a few swiftly exchanged terms, old language is revivified: according to Karl Lashley (1950/1988: 59), a mere substitution of nerve impulse for animal spirit, and of synapse for brain pore results in 'the doctrine of learning as change in resistance of synapses'.[1] Now computational neuroscience, announces Patricia Churchland in *Scientific American* (1989: 100), is modernising Descartes' vision of reflex action in neural nets.

Obviously, such heroic nostalgia answers no better to historical or philosophical complexity than does easy vilification of Descartes as source of 'the original sin of modern philosophy' (Rorty 1980: 60). Canguilhem (1955/1977: 138–42 (= 1994: 52–6)) marvellously undermined rhetoric of Descartes' 'paternity' by analysing the specific political circumstances in which Descartes was 'anointed' an illustrious precursor for 'discovering' the reflex by du Bois-Reymond in the 1850s. In such myths, invocation of the oddly absent father (as hero or villain) distorts and distracts from important current polemics (Tomaselli 1984a; Williams 1994).

Safety, then, requires historians to avoid compressing old and new, wary of the dangers of reducing an alien past to the present. But I have a different agenda, in which a skewed historical slant on one topic, memory, can open conceptual space in modern sciences of mind. I am interested neither (though a materialist) in another anti-dualist exposure of 'Descartes' error', nor (though agreeing with Churchland that neural networks are not just new fashion) in identifying Descartes as exuberant neurophilosopher. My retrospective reading, on the basis of current concerns about memory, is none the less situated within a growing body of work on Descartes' natural philosophy which seeks to counteract historiographical obsession with metaphysics (and I extend to memory the new approaches to internal sense developed by Baker and Morris (1996) and by Sepper (1996), whose books appeared too late for me to use them fully). Anti-naturalist history either neglected Descartes' science or treated it only as an adjunct to first philosophy (Kenny 1968; Frankfurt 1970; Williams 1978; Wilson 1978; Cottingham 1986; Dicker 1992). Neither the Descartes of undergraduate courses nor that of psychoanalytic and anti-humanist polemic answers to the more interesting, weirder texts in natural philosophy (Irigaray 1974/1985a and Kofman 1976/1991, for example, focus on metaphysics and the *Meditations* in particular).

What happens if we take seriously Descartes' advice that meditation on metaphysical principles is 'very harmful' since it impedes us from imagination

1 Descartes' memory traces are Semon's engrams (Riese 1958: 125–6); his 'system of tubes' can be replaced by 'the present system of concatenated neurons' (Foster 1901/1970: 268, in Carter 1983: 19).

and the senses (to Princess Elizabeth, 28 June 1643, AT iii.695, CSM-K 228), and note that about 90 per cent of his surviving correspondence is on scientific matters (Gaukroger 1993a)? Descartes did not set out implausibly to derive the whole of physics from the *cogito*: early modern followers and critics alike treated Cartesianism as over-hypothetical and speculative, rather than a prior-ist and deductivist. Descartes' place as anti-hero in neat schematic battles between rationalism and empiricism was imposed by nineteenth-century historians (Gaukroger 1993a, 1995: 5–6): earlier readers were as likely to study the psychophysiology of *L'Homme* as the *Meditations* (Wright 1983: 5–9). In describing the neurophilosophical model of memory, I assume Descartes' interest in knowledge gained a posteriori (Larmore 1980). On this reading Descartes was a mechanical philosopher who was as interested in the motions of matter as in the supernatural realm carefully separated from it (Hutchison 1983; Schuster 1990; Tamny 1990): although a dualist, he was uninterested in metaphysical dualism (MacIntosh 1983: 328; Clarke 1982).

It is not that Descartes was a pure empiricist either. Although the popular view that 'Descartes shunned experiment with Jesuitical disdain' (Miller 1978: 295) is quite wrong in *general*, the physiology seems particularly ridiculous just because of the gap between accounts of inner fluid machinations and any conceivable observation. There is still disagreement on how Descartes *did* see the relations between metaphysics and scientific practice (Hatfield 1985, 1990: 111–17; Gaukroger 1993a; Schuster 1993); and scholars still convict Descartes of inconsistency for not adhering firmly in practice to his methodological 'order of reasons'. His speculations on the physiology of memory are a useful test case in assessing the significance of Descartes' own neglect of Cartesian method.

Descartes outlined the rudiments of a distributed model of memory, where memories are 'stored' only superpositionally. 'Memory' is neither a repository for images nor an inner lexicon. Distributed representation can be theorised at a level of abstraction from the specific (historically and technologically dependent) material in which it is thought to be implemented. Descartes was not the origin of such a model: the key questions about interference and mixture in memory which these models bring up were discussed in the long traditions of the memory arts (Sutton 1997). But Descartes was hostile to the cognitive technology of place-memory systems (Sepper 1993; Gaukroger 1995: 273–4), rejecting humanist schemes for the classification and rearrangement of already existing knowledge (Judowitz 1988: 25–32; Gaukroger 1989: 31–8, 46ff.). Architectural mnemonics (Carruthers 1990) was designed to supplement weak, confused natural memory with rules for maintaining rigorous order. Descartes instead examines sources of this 'natural' confusion in memory. The psychophysiological ideas he draws on were not intended to locate memory's seat in the brain, but to model the mechanisms of retention

and storage. To the extent that his neurophilosophy is absurd, it is partly because of his hidden reliance on sources with which we have no illusion of familiarity. But he takes seriously as explananda holistic associative processes, phenomena of blending and interference, and the uncertain relation of memory and imagination.

This reappropriation of Descartes is the deliberate antithesis of uses of Descartes-the-rationalist by Chomsky and Fodor to buttress the language of thought hypothesis and 'Cartesian linguistics' (Chomsky 1966). Rather than taking the defining feature of Cartesianism to be the stress on rationality, linguistic generativity, and innate ideas, I point underneath to Descartes' (much more extensive) work on the many capacities which we share with other animals. What Fodor denigrates as babble and mere mental causation (chapter 11 below) is as central to Descartes' concerns as deduction, intuition, or meditation.

I begin by outlining evidence for distributed representation in Descartes (section 3.2), and develop the theme in answering a series of objections to my reading. The objections are powerful, and will occur to readers immediately, so I mention them here at the outset before addressing them in turn in sections 3.3 to 3.6.

1 Intellectual memory: notoriously, Descartes believed in an incorporeal intellectual memory: does this not vitiate my stress on his physiology of memory?
2 Memory, soul, and automatism: notoriously, Descartes denied that animals have souls: does this not vitiate my assumption that his physiological works alone, describing soul-less machines, could even address problems of memory and cognition at all? How can a dualist have a neuroscientific theory of memory? More specifically, behaviour without the soul as cause is for Descartes, notoriously, purely automatic, the mere product of reflex arcs: does his restriction of physical behaviour to reflex action (accepted even in Churchland's search for precursors to connectionism) not vitiate my claim that complex and flexible action occurs even when the soul is not involved?
3 Mechanism, matter, and active bodies: notoriously, Descartes' mechanical philosophy requires that matter (including the human body) be passive, pure extension in motion: does this not vitiate my picture of a dynamic physiology in which the body is always active, always escaping the soul's command? Is mechanism not an intrinsically authoritarian natural philosophy, in which isolated atomic elements combine only through external intervention, rather than the anarchic holism of my account? Is the Cartesian body not inert, closed, a forgotten container, rather than the highly theorised, permeable temporary pocket of stability embedded in social and physical worlds which I attribute to Descartes?

4 Memory, method, and metaphysics: notoriously, Cartesian method requires the subject to doubt opinion and easy belief in starting anew with only clear and distinct ideas. As Tim Reiss (1996) shows, Descartes does not claim (as Gassendi thought he did) that old ideas are simply erased, obliterated in the process of destruction and rebuilding: is he not then acknowledging the permanence of stored items in a way which vitiates my claim that memories, for Descartes, must be reconstructed?

These issues overlap in their relevance to memory but together touch an enormous range of Cartesian problems. My answers come at a small set of claims from different perspectives, trying to get across unusual points of view about Cartesian automata, about Descartes' conception of the body, and about the place of the past in Cartesian method.

Even if my responses to these objections are adequate, Descartes remains not just historically but conceptually distant from modern distributed-memory theorists. Unable to integrate his psychophysiology of memory into his general philosophy, Descartes is least clear and least convincing on issues of cognitive control: not just about familiar problems of the non-physical soul, but on the independent issue of the need for an executive centre in the brain, whether hooked up to a soul or not (Dennett 1991a). I conclude by briefly drawing together the various threats to psychological control and mental autonomy, challenging 'Descartes' relentless optimism about the autonomy and power of the will' (Rorty 1992: 384), which are posed by memory and the body, and sketching a more positive Cartesian account of self-mastery based on moral neurology.

In two appendices I fill in background to this Cartesian philosophy of the brain. The first spells out the general physiological psychology of *L'Homme* to show how the theory of memory is embedded in it. The second examines Malebranche's more explicit distributed model of memory, which supports the attribution of distributed representation to Descartes, as does the horror expressed at its consequences by English critics (chapter 5 below). In the late seventeenth and early eighteenth centuries, both supporters and critics of Cartesianism were aware of, and concerned about, the implications and dangers of these views of memory and brain. These accounts are unfamiliar now, hidden under the weight of commentary on scepticism, the *cogito*, and foundational epistemology; but it is naive to ascribe this to progress in separating truly philosophical questions about knowledge and mind from truly empirical questions about brain traces.

This long chapter cuts new paths through many strands of Descartes scholarship, at the cost of starting from a highly specific concern with memory. The neglected memory models map on to slants on surrounding issues that deserve attention. Relations between body and world, between psychology and physiology, between forgetting and moving on are at stake in memory, and

Descartes hinted at much more interesting views on all these topics than our textbooks admit.

3.2 Distributed representation in Descartes' neurophilosophy of memory

Remembering, for Descartes, is (or at least requires) the reconstructing of patterns of motion in the animal spirits flowing through particular brain pores. This reconstruction is possible in retrieval because of physical dispositions in these pores: these dispositions or memory traces are *superpositional* in that there can be many in one part or fold of the brain. He describes associative mechanisms which allow not only for redundant coding and graceful degradation, but also for causal holism in the effects of implicit memory representations. They chime with his interest in external forms of condensed representation, such as the distorted images of anamorphotic art which required reconstruction from specific points of view (Decyk forthcoming). I do not deal with the aesthetics of superposition, and here merely introduce the model of memory to exemplify distributed representation.

L'Homme

In April 1630, soon after starting anatomical studies, Descartes told Mersenne that he was studying chemistry and anatomy simultaneously: 'every day I learn something that I cannot find in any book' (AT i.137, CSM-K 21; compare to Mersenne, 18 December 1629, AT i.102, and 20 February 1639, AT ii.525, CSM-K 134). In late 1632 he was 'dissecting the heads of various animals, so that [he could] explain what imagination, memory, etc. consist in' (AT i.263, CSM-K 40).[2] Reading and observation resulted in revisions of his earlier views on sensation and cognition.[3] *L'Homme* (the *Treatise*

2 The claim to be discovering facts unavailable in any book should not be taken at face value; in fact Descartes told Mersenne in 1637 that his anatomical assumptions were those accepted generally (AT i.378). Hall's commentary details Descartes' sources. But here as elsewhere, in Alquié's words (1966: 27), information read by Descartes entered 'une mémoire déformant ce qu'elle rappelle, et toujour mêlée à la création'. This lovely description of Descartes' own memory also applies to his theoretical conception of memory. It is quoted by Annie Bitbol-Hespériès, who throws new light on Descartes' physiological reading, demonstrating the particular importance of the early seventeenth-century work of Caspar Bauhin (1990: 36, 195–202, 214–18). Not much of her new material relates directly to memory and other internal senses, but I hope it is clear that my narrative, playing with a new myth of Descartes' 'discovery' of distributed representation, skates over a more complex past. Provocative accounts of Descartes' anatomical contacts and practices in Leiden and Amsterdam include Lindeboom 1979: ch. 3; Barker 1984: 73–85; Rupp 1990, 1992; and Sawday 1995: 146–58.

3 I cannot address here the account of cognition in the earlier *Regulae*, where Descartes assumes that objects stamp imprints on sense-organs which are instantaneously transmitted through the *sensus communis* into *phantasia*, where they are recognised by the *vis cognoscens*. On this psychophysiology see Sepper 1996: 28–35 (compare Sepper 1988, 1989,

on Man)[4] resulted. Though it was published only in the 1660s (Descartes 1662, 1664), it was almost complete by July 1633, when, on hearing of the condemnation of Galileo, Descartes prudently decided not to risk antagonising the Church by publishing *Le Monde*, of which it was part. I give some background to *L'Homme*'s strange general physiology in appendix 1 to this chapter, but here move straight to memory.

L'Homme describes a soul-less world in which earthen machines imitate our bodily functions. But these 'statues' are animated, these machines dream. The capacities of the self-moving automata are many, though they depend only on the disposition of the organs (AT xi.120, H 4, CSM 1.99; AT xi.201–2, H 113, CSM 1.108):[5] there is no life/body dualism to accompany Descartes' mind/body dualism (Mackenzie 1975; Wright 1990: 253–4). Cartesian cyborgs can not only walk, breath, sleep and wake, nourish themselves, digest, and reproduce: they also have what are *to us* mental capacities like sensation, imagination, memory, and emotion. Descartes' fable seeks to catch at the very pulse of the machine.

All this is accomplished (thanks to God's skilled craft)[6] by means of the whooshing animal spirits, shaking through brain tissues. The spirits incessantly undergo *criblage* or *tamisage* (sifting, filtering, sieving) in the textured porous net, forming and retracing patterns across the inner surfaces of a filamentous mesh. They connect the deepest interior, the pineal gland, to the world in numerous ways. It is not just that muscular motion requires appropriate spirituous input from brain to periphery. Rather, interior and world are always already connected, for the state of the body at a time results from the history of its interactions. The condition of animal spirits shifts with context: changes in environment, climate, diet, bodily practices, in the condition of the

1993); Foti 1986b: 634–7; Gaukroger 1995: 158–72. These writers do not draw attention to the absence of animal spirits in the *Regulae*: without the spirits, no dynamics of memory could be even implicit.

4 The gender-specificity of the subject of Descartes' inquiry is uncertain. Where Stephen Voss (1994: 273) translates Descartes' 'unhappy usage' as 'man', I retain the French, and hope to show that the gendered body-machines in Descartes' overtly universalising text are in fact all particular, context-bound, and marked by bodily differences. I do not address Descartes' views of gender explicitly here: but my revised interpretations of bodies and memory may help such projects. See also Sutton 1998, and chapter 9 below on how animal spirits focus Cartesian men's fears that psychological and libidinal energies were mutually exclusive.

5 For *L'Homme*, in addition to the editions cited, I have used Alquié's edition and notes (A 379–480), the Clark Library copies of the first Latin and French editions (Des Cartes 1662, Descartes 1664), and an unpublished translation of *Le Monde* and *L'Homme* by Stephen Gaukroger.

6 The teleology in the relation between the fable's machine world and its real-world counterpart is double. Firstly, God benevolently ensures mapping between material processes and useful outcomes. But, secondly, what counts as useful, what is conducive to life, is given only biologically. This, as Canguilhem points out (1952, 1955/1977: 31–2, 54–6), is not necessarily a lapse from mechanism, since mechanisms have functions (compare A 403, 410, 429; Hatfield 1992: 361). But finality is absent in *L'Homme*: the creator, like the author, disappears behind the fable.

blood and other body fluids, change the spirits, whose influence on cognition in turn gives rise to ongoing changes of state for the body in the world. This is a field of multiple simultaneous interactions, akin to our understanding of dynamical systems in which 'everything is simultaneously affecting everything else': the system's state variables and the external parameters which influence their evolution are all changing in various timescales (van Gelder and Port 1995: 9, 23–5; compare Clark 1997: 163–6).

In focusing on the spirits' roles in memory, I bypass the separate 'intellectual memory' until section 3.3 below. A deliberately partial reading makes good sense: both followers and critics fixed only on the neurophilosophical distributed model. Making this model clear poses questions about relations between the two kinds of memory, and about motivations for retreating from the corporeal account. Hall comments (H xxxix) that Descartes 'gives a suggestive model for the hypostasis of memory, wrong in its details but right in its assumption that a physical basis of retention must exist', without saying which details are wrong. Morris (1969: 460), characterising the model only as 'mechanical', mentions Descartes' 'constant retreat away from any commitment to the model as a total explanation of human memory', as he relied increasingly on non-physical intellectual memory. Though Descartes clearly was not satisfied with his corporeal model, understanding of its nature and implications requires a different theoretical framework. One commentator acknowledges the reconstructive nature of memory in Descartes, but wrongly assumes that only incorporeal memory could be reconstructive, on the grounds that corporeal memory must be passive and static (Landormy 1902: 283–7). I deny this.

Corporeal memory in L'Homme

In the section of *L'Homme* on memory, ideas are explicitly defined as impressions or figures which animal spirits trace on the surface of the pineal gland as they leave it (AT xi.177, H 86–7, CSM 1.106). These idea-impressions derive either directly from sensory impressions or from imagination and several other (internal) causes: contact with reality is possible, but uncertain in any particular case. Whatever their source, traces of these idea-patterns are carried by spirits flowing from the gland and 'imprinted in the internal part of the brain, which is the seat of Memory' (AT xi. 177, H 87, CSM 1.106). In the accompanying diagram (Figure 1), the part referred to appears as the large, relatively undifferentiated fibrous mesh of the brain substance. How does this imprinting work?

Declining to comment, as he says he could, on how these traces 'can sometimes even be caused, by certain actions of the mother, to be imprinted on the limbs of the child being formed in her entrails' (AT xi.177, H 87, CSM 1.106; chapter 9 below), Descartes sketches a theory of recall or retrieval. Animal spirits leaving the gland move through the ventricles towards different regions of the brain substance. The spirits pass, Descartes continues,

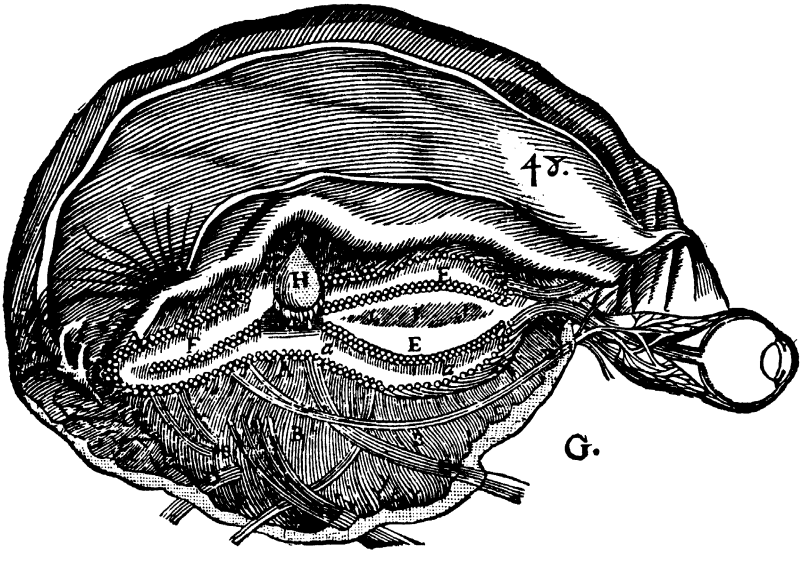

Figure 1 The memory system in Descartes' *L'Homme* (Descartes 1664: 74)

> into the pores or intervals that occur between the filaments composing part B [the solid part] of the brain. And [assume] that they are forceful enough to enlarge these intervals somewhat and to bend and rearrange [*plier et disposer diversement*] any filaments they encounter [*en leurs chemins*], according to the differing modes of movement of the spirits themselves and the differing degrees of openness of the tubes into which they pass. *Thus they also trace figures in these gaps, which correspond to those of the objects.* [Assume also] that the first time they accomplish this they do so less easily and effectively here than on gland H, but that they accomplish it increasingly effectively in the measure that their action is stronger, or lasts longer, or is more often repeated. Which is why in such cases these patterns are no longer so easily erased, but are retained there in such a way that by means of them the ideas that existed previously on this gland can be formed again long afterward, without requiring the presence of the objects to which they correspond. And it is in this that Memory consists. (AT xi.178, H 87–8, CSM 1.107)[7]

The pattern of the pores, which constrains the patterned flow of spirits, is itself altered over time by the differing motions of the spirits. Note first that patterns are not *stored* faithfully, to endure separately one from another: they are just retained 'in such a way that' they play a part in the (re-)creation of the idea on the surface of the gland. This is representation without resemblance, suggesting a dispositional model: patterns are 'stored' only implicitly, not kept

7 The emphasised sentence (Des Cartes 1662: 83, Descartes 1664: 74) is omitted by Hall, and I use Stoothoff's translation in CSM. The insertions in square brackets are Hall's, except for phrases from the French original which I give to confirm the sense.

in exactly the same form throughout the interval between experience and remembering.

Note too that reconstruction depends only upon physical factors. Spirits fold, enlarge, bend, and arrange the relevant fibres of the brain substance into forms which will in turn filter and sift the spirits' future flows. The critical variables are the degree and pattern of openness of the passageways in the fibrous substance, and the direction and strength of the flow of spirits. The whole scheme operates by contact action alone, and yet patterns, which 'correspond' to absent objects, can be retained as tendencies or dispositions for their reconstruction or re-evocation. The soul may be involved, when it is attached to the machine: but it is not necessary for memory operation.

In this associative mechanism, repetition of recall is one aid to easy recall in the future. Its physical realisation falls out of the model: after spirits have entered the same passageways more and more often, they will tend to find them again more quickly, to enter with greater force, and to remain longer before leaving. Even if particular passageways (through the brain pores) which need to be open for the reconstruction of a particular pattern happen to be closed over, they still 'leave a certain arrangement [*disposition*] of the filaments composing this part of the brain by which they can be opened more easily later than if they had not been opened before' (AT xi.178, H 89). The microstructure of the brain, altered in the course of experience, makes the spirits' future access to pores easier and quicker.[8]

Descartes gives an analogy with a linen cloth (*une toile*) which has had 'several needles or engravers' points' repeatedly passed through it (Figure 2). Some holes in the cloth will remain open after the needles have been withdrawn, but even if they close, physical 'traces' left in the cloth will enable them to open again easily (AT xi.178–9, H 89). This mechanism allows total recall on partial input ('the recollection of one thing can be excited by that of another which was imprinted in the memory at the same time'): if only some holes are reopened, this may still suffice for the simultaneous reopening of the others, 'especially if they had all been opened several times together and had not customarily been opened separately' (AT xi.179, H 90). Likewise, 'if I see two eyes with a nose, I at once imagine a forehead and a mouth and all the other parts of a face, because I am unaccustomed to seeing the former without the latter', and seeing a fire I remember heat.[9] An incomplete input can thus generate or prompt a complete

8 As Hall says (H 96, n. 145): 'memory traces . . . consist in residual patterns of openness among the interstices of the filamentous brain substance'.

9 Morris (1969: 454) feels that this is not legitimately mechanical because of the mentalistic content of the term 'accustomed'. But this is merely the phenomenological explanandum, Descartes' common-sense illustration of a phenomenon which his theory explains: the actual theory has already been given, and relies only on a repeated pattern leaving physical effects, the nature of which change over time. Morris is bewitched by 'mechanism', assuming it must exclude all temporal patterning.

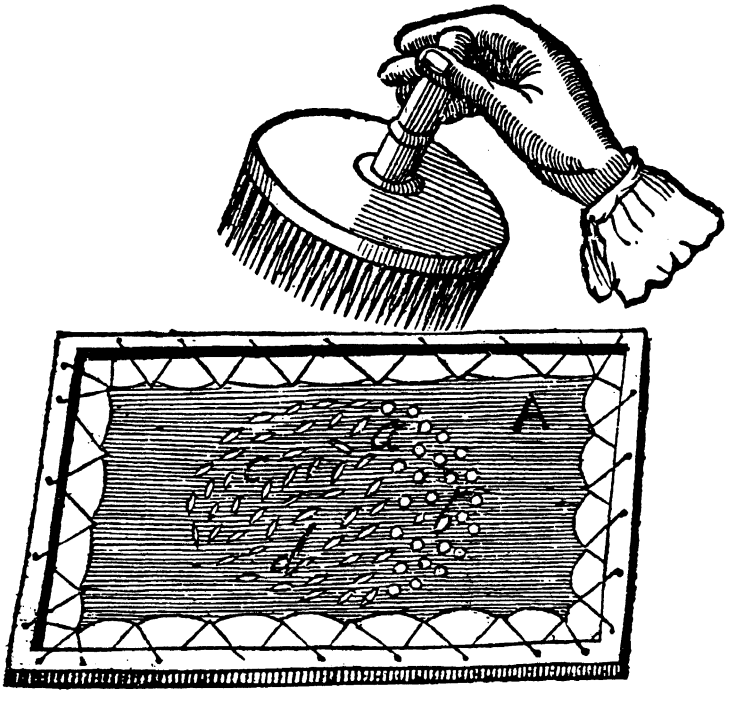

Figure 2 Distributed patterns in a linen cloth as an analogy for associative memory (Descartes 1664: 76)

memory. Memory representation is a form or associative extension of natural indication (Mackenzie 1992: 133–41).

In the *Passions of the Soul* (1649), Descartes all but repeats the account (*Passions* 1.42, AT xi.360, CSM 1.344):

> These traces consist simply in the fact that the pores of the brain through which the spirits previously made their way owing to the presence of this object have thereby become more apt than the others to be opened in the same way when the spirits again flow towards them. And so the spirits enter into these pores more easily when they come upon them.

But none of this hints at more precise mechanisms for retention and storage of these traces. What is the state of brain pores and animal spirit motions during retention, when a particular trace is unactivated, which allows its future reconstruction?

A distributed model of memory

Descartes had at least an inchoate understanding of *superpositional storage*, the defining feature of a distributed model (van Gelder 1991b). While the main

evidence that Descartes saw this possibility is in correspondence, there were firm hints in a less familiar part of *L'Homme*. After discussing memory, Descartes describes two corporeal causes of the pineal gland's movements. It can be affected by 'the action of objects that impinge on the senses' (AT xi.185–8, H 96–100). But Descartes gives extended attention to the influence of 'differences among the particles of the spirits that leave' the gland, concluding with an expansion of the theory of memory (AT xi.180–5, H 91–6).

After explaining the internal dynamics of the whole system of gland, brain tubules, animal spirits, nerves, muscles, and limbs, Descartes acknowledges that the pineal gland is often inclined in particular directions by the force of the animal spirits alone, without the involvement of rational soul or external senses. In such cases,

> the ideas that are formed on its [the gland's] surface proceed not only from inequalities in the particles of the spirits causing corresponding differences in temper [*des humeurs*], as mentioned before, but also from the imprints of memory. For if at the region of the brain toward which the gland is inclined, the shape of one particular object is imprinted more distinctly than that of any other, the spirits tending to that region cannot fail to receive an impression thereof. And it is thus that past things sometimes return to thought as if by chance [*comme par hazard*] and without the memory of them being excited by any object impinging on the senses. (AT xi.184, H 96)

This is not the mere automatism of a reflex model, for internal states and processes are maintained and reactivated, and have causal effects, with some autonomy from present sensory stimuli: this is the source of moral danger. The passage implies that there can be more than one trace 'imprinted' in the same 'region' (*endroit*) of the brain. Due to the history of the machine and to the (physical) context, spirits pass through interstitial tubules in one particular pattern, thus reconstructing one distinct trace of all the traces there stored. None of this *requires* a soul: *comme par hazard* means 'without control' as much as 'by chance'. Memory is the scene of a triple displacement of control: neither the non-physical soul, nor the pineal in its role as central executive, nor external reality, *need* inevitably be the cause of ideas. Even when these three putative guarantors of rationality are involved, the decentring forces of superposed memory motions may still influence the context-bound formation of ideas.

Traces are distinct only when reconstructed. But it 'usually happens' (*le plus souvent*) that 'several different figures are traced in this same region of the brain almost equally perfectly': they do not retain independence in this implicit state. When this occurs,

> the spirits will acquire a [combined] impression of them all [*quelque chose de l'impression de chacune*], this happening to a greater or lesser degree according to the ways in which parts of the figures fit together [*selon la diverse rencontre de leurs parties*]. It is thus that chimeras and hypogryphs are formed in the

> imaginations of those who daydream, that is to say who let their fancy wander listlessly here and there without external objects diverting it and without the fancy's being directed by reason. (AT xi.185, H 96)[10]

Memories are motions, not bodies, and animal spirits, like other exquisitely small particles, can receive and transmit many different motions at the same time.[11] With traces already mixed, it is all too easy for them to interfere one with another: when spirits acquire combined impressions, the implicit traces' causal effects are holistic, everything in a region making a difference in ongoing processing. There is no fixed *place* for a single 'memory'. As every sensation is, in a sense, many sensations, so every Cartesian memory is many memories: the prejudices of the past are always with us, always to be detected.

Imagination here is the work of memory rather than a separate capacity, a disturbing possibility which haunts the moral physiology of the following century. Although, to stress the moral dangers of brain processes, Descartes uses fantastical examples, suggesting 'the imagination's freedom from natural occurrence, its ability to rearrange wings and legs into new combinations, and . . . [its] eventual capacity to create beyond "need"' (Scarry 1985: 163), the same mechanism operates in ordinary imagining and remembering. There are differences only of degree in the confusion of parts in figures which spirits trace: freedom from melding is either rare or illusory. 'Those who daydream' cannot be securely cordoned off from rational thinkers, who are also subject to the combination and condensation of traces.

In 'Descartes' assimilation of imagination and memory' (Foti 1986b: 636), the direction of explanation is uncertain. Is it that cognitive theory produces a possibility surprising to common sense and to morality, or that a certain part of common-sense psychology drives or suggests the way the cognitive theory is developed? This ambiguity, which recurs in modern relations between cognitive and cultural theory, needs no resolution: scientific and folk conceptions coevolve. Alquié complains (A 453) that, when dealing with associative memory in *L'Homme*, Descartes transposes to the organic realm ideas which the psychological realm had supplied, and that to this extent, 'he ceases explaining the psychological by the physical' (compare Grosholz 1991: 117–30). But sciences of the organic and the physiological need explananda from somewhere: looking at a few brains is not enough to render salient the key questions for neuroscience. It is not sufficient criticism of the terms of a physiological theory

10 'Combined' is Hall's elucidation. Landormy (1902: 280–1) summarises nicely the role of memory traces in the mechanics: 'les esprits . . . peuvent faire renaître . . . des débris de diverses images autrefois perçues et les combiner en des ensembles imprévus'. In a marvellous, mocking exposition, Krell (1990: 72–3) notes that impressions absorbed 'higgledy-piggledy' are 'prone to moral turpitude, lassitude, lethargy, and benumbment'. These concerns recur in subsequent debates on distributed memory: see chapters 5 and 9.

11 The analogy here is with interference phenomena in the theory of light. See *Le Monde* ch. 14, AT xi.101; Sabra 1967/1981: 17–68; Shea 1991: 212ff.; chapter 4 below.

that they derive from psychological metaphors: only overly pure ideals of science require the two levels connected in reductive explanation to have had no prior metaphorical contact.

Descartes talks of imprinting, of figures and shapes: but it is obvious that these are not resembling images of dubious ontological status. 'Figures' are (transient) patterns of openness of brain pores, and the 'parts' of particular figures are the specific tubules which are jointly involved in the reconstruction of the whole figure corresponding to an absent object. But Descartes says no more about what happens to particular traces while unactivated, between their involvement in separate episodes of remembering, until a sequence of letters in 1640.

Impressions preserved in memory, Descartes firstly suggests, are 'not unlike the folds which remain in this paper after it has once been folded' (to Meysonnier, 29 January 1640, AT iii.20, CSM-K 143). Memories are retained in different areas of the brain: impressions are 'received for the most part in the whole substance of the brain', as well as on the gland itself. But this is not yet superposition, for traces extended across wide regions could still be independent of each other (van Gelder 1991b: 48–53).

In the letter which first introduces an intellectual memory, which 'depends on the soul alone' (to Mersenne, 1 April 1640, AT iii.48, CSM-K 146), Descartes is still primarily concerned with specific physical locations for memory impressions, whether in the gland, in the brain, in the nerves and muscles, in the hands of a lute player, or in external aids such as on the pages of a book. The body and the book both support the brain's memory capacity which, because prone to confusion, is fragile. Boundaries blur between 'natural' and artificial memory systems. It is not that brains and books do their recording in the same way, as language of thought theorists think (Fodor 1976): it is that the symbiotic interaction between brains and books is necessary because distinct recording is not common in brains, which thus require certain forms of supplement.

Morris (1969: 455) takes these suggestions about external storage to show that 'Descartes was having trouble finding room in the brain for all the images that his purely physical theory required': he was driven to 'resort to the doctrine of an "intellectual memory"' on finally realising the insufficiency of storage space in the brain. But this is not the correct diagnosis. The problem of finding room in the brain for all of a person's memories was debated in the seventeenth century (MacIntosh 1983). Descartes had worried about it before adopting an animal spirits theory of memory: in the *Regulae*, he identified the phantasy ('a genuine part of the body') with memory in that it 'is large enough to let the different parts of it take on many different figures and generally to retain them for some time' (Rule 12, AT x.415, CSM 1.41–2). Representations had to be independently stamped into the memory wax: this is why the problem of room

in the brain arose. But by 1640, Descartes could deflect the problem, since he allowed more than one trace in the same place.

Morris overstates Descartes' concern: in the same letter in which intellectual memory is introduced, Descartes confirms that, despite reference to external aids, it is 'especially the interior parts' of the brain 'which are for the most part utilized in memory' (CSM-K 146). Two months later Descartes elaborates on *L'Homme*'s account of interference between memory traces:

> There is no doubt that the folds of the memory get in each other's way, and that there cannot be an infinite number of such folds in the brain; but there are still quite a number of them there. Moreover, the intellectual memory has its own separate impressions, which do not depend in any way on these folds. So I do not believe that the number of folds is necessarily very large. (To Mersenne, 11 June 1640, AT iii.84–5, CSM-K 148)

Memory impressions are not now *like* folds in paper: there are real '*plis de mémoire*', folds of memory. If traces are in the changing patterns of brain pores which guide spirit motions, these patterns 'get in each other's way' by interfering in the processes of retrieval. But it is *because* different traces are retained in the *same* parts of the system that only a finite number of such parts (the folds) is required.

This is confirmed when Descartes replies to Mersenne's query about the numerical sufficiency of the brain's 'folds' for all our memories: 'I do not think that there has to be a very large number of these folds to supply all the things we remember, because a single fold will do for all the things which resemble each other' (to Mersenne, 6 August 1640, AT iii.143, CSM-K 151).[12] Many memories can be stored superpositionally in the same system. The particular memory elicited from the system at a time depends on the relation between new patterns of animal spirits entering the folds and the existing dispositional states of spirits and brain fibres.

Descartes refers again here to an 'altogether spiritual' memory, which 'is not found in animals', but which 'we mainly use' (CSM-K 151). The fact that intellectual memory is suddenly given a central billing just as Descartes is spelling out more details of the distributed model is obviously significant. I ascribe his increasing need to back up the account of corporeal memory with a catch-all non-physical memory to an awareness that distributed models led to the interference which, in *L'Homme*, was already pinpointed as a source of daydream, fantasy, and moral slackness. But the new stress on intellectual memory leads

12 This notion of 'resemblance' needs spelling out: Descartes does not think traces represent objects by picturing them (chapter 15 below). In the next sentence Descartes wrote of the bodily memory's 'images', but scored this word out in preference for 'impressions', perhaps making clearer the non-resemblance between patterns of motions in the brain and the objects of memory.

Morris (1969: 455–7) to the stronger conclusion that Descartes has *abandoned* the earlier theory by relegating purely physical memory to beasts alone. But, to repeat, Descartes gives the same account of memory in the *Passions of the Soul*, his last published work, as in *L'Homme*. There are detailed parallels between the psychophysiological theories of the two works (as in the account of hatred): but there is no mention in the *Passions* of intellectual memory. Morris has to downplay the later text, making the strange suggestion that 'Descartes was not yet prepared to defend the doctrine [of intellectual memory] in public' (1969: 457). This is implausible, given the doctrine's theological orthodoxy and respectable position in scholastic and Aristotelian philosophy (Kessler 1988: 509–18).

Descartes' references to 'folds' of memory are compatible with the mechanical theory of *L'Homme*. Traces which remain in the brain after images are imprinted dispose the brain 'to move the soul in the same way as it moved it before, and thus to make it remember something. It is rather as the folds in a piece of paper or cloth make it easier to fold again in that way than it would be if it had never been so folded before' (to Mesland, 2 May 1644, AT iv.114, CSM-K 233). Here the metaphor of the folds is used in exactly the same way as that of the holes left by needles in a cloth had been in *L'Homme*, to explain the mechanism of association, whereby repetition of recall renders further recall easier still. So Descartes adumbrated, albeit hesitantly, a distributed model of memory which does not require discrete, independent storage of all memories and their possible combinations, and which was taken up by later supporters and critics.

The model also implies peculiar relations between processing and storage. Strictly speaking, ideas only exist on the surface of the pineal gland. But memory serves to complicate clean divisions between insides and outsides. There are also superpositional memory representations *outside* the pineal, in the folds of memory throughout the internal substance of the brain and in the body too. In fact memory impressions will only be predominantly located on the pineal gland in people whose minds are 'coarse and sluggish': clear thinking requires the continual externalisation or projection of memory traces out from the gland into supplementary storage systems, in brain, nerves, muscles, hands, and on paper (to Meysonnier, AT iii.20, CSM-K 143; to Mersenne, AT iii.48, CSM-K 146). This is, again, not because Descartes is worried about finding enough memory room in the brain, but so that the pineal remains open to the tracing of new figures. This need can never be fully satisfied, since spirits move continually through the gland from the memory folds. Animal spirits are always breaching the resistant memory folds which are in turn modified in a continual play of difference (compare Foti 1986a: 78).[13]

13 Foti (1986a: 76) quotes Derrida's (1978b: 200) description of Freud's problematic of 'accounting simultaneously for the permanence of the trace and for the virginity of the receiving substance'. She argues that Descartes is banishing or relegating

It is tempting to assume that even those actions in which the soul is not involved are still driven by a single central executive where 'it all comes together' (Dennett 1991a: 107; Dennett and Kinsbourne 1992). Dennett's sustained attack on the persisting metaphor of the centre is meant to force acceptance that, once dualism is abandoned, it is still necessary to undermine the further intuition that there must be a headquarters inside the brain, an inner sanctum or functional centre where input must terminate and be received, and output be initiated (1991a: 104–8, 144, 257–8, 321–2). No executive nerve centre exists, for cognition and control are 'distributed around the brain' (1991a: 169).

Dennett's choice of Descartes as target here appears natural, for the pineal gland, unique in its unified structure (*Passions* 1.32, AT xi.352–3, CSM 1.340), seems uniquely to control all processing.[14] But this is not so, and Dennett's label 'Cartesian materialism' in fact applies poorly to Descartes' own views. The kind of 'processing' accomplished by or at the pineal gland is minimal. Ongoing processing actually occurs out in the interactions between input and representations in the folds of memory, in the same system as storage. No entities or independent items are lifted from memory boxes, transported to the gland to influence computation, and then dumped back in their places: that picture of executive control only works for local models with separate fixed items, as I will describe in Hooke (chapter 5). There is space for transcendent central control in the Cartesian philosophy of the brain: but it is very limited in scope and frequency.

confusion, 'errant memory inscription', and difference 'to the outside, beyond the boundaries of presence', to preserve 'the ideal virginity of the present'. But my reading differs, as I do not ascribe to Descartes a belief that memory *has* a 'proper inscription in the interior parts of the brain' (Foti 1986a: 77–9). In the Cartesian distributed model, as in Derrida's reading of Freud, both desiderata of the problematic are denied: traces are nowhere permanent, and neither pineal gland nor brain is ever clean.

14 Whatever the best interpretation of its role, the discussion of the gland in the *Passions* is not a joke (Grene 1985: 48) but an edited, matching version of that in *L'Homme*, and it places the mental firmly in the causal field. On the pineal gland see Lindeboom 1979: 81–3; Krell 1988, 1990: 319–20. There is a suggestive new theory about its (still poorly understood) function: Maestroni and Conti (1991: thanks to Doris McIlwain for this reference) suggest that it has a central role in 'the psycho–neuroendocrine–immune network', in regulating fertility and environmental rhythms (1991: 496), and (for humans) in causing 'affective and psychosomatic diseases, psychiatric and neurological disorders, and cancer' (1991: 496). This role is mediated by the nearest modern analogue to animal spirits, neuropeptides, and if disturbed may cause differential perceptions of and responses to stress (1991: 510–11). Ann Mackenzie (1989: 175) makes the tempting point that the 'general bodily factors' which Descartes accepts as causal factors in cognitive processing include 'some . . . which we would view as importantly hormonal'. Animal spirits, like hormones, were 'primarily conceptualized in relation to notions of control, encapsulating possibilities for both loss of control, through hormonal lack or excess, and (medical) interventions to (re)assert control over multiple and minute elements of life processes' (Harding 1996: 100).

3.3 Objection 1: intellectual memory

Surely I cannot be serious: Descartes the connectionist? No. Yet the concept of distributed memory is independent of the specific technological and computational developments of recent years. Many apparent reasons to resist my interpretation of the Cartesian philosophy of the brain turn out not to hold up. In this and the next three sections I run through a set of objections to my account. In answering them I add more detail to the positive characterisation of the models of memory in which I am interested. Descartes is not always, and never clearly, close to an idealised historical connectionist: I seek the vanishing points in his view of memory, where he holds back from consequences of his own theory. The dialectic in these sections is twofold. I show, in each case, that the objection in question does not unambiguously refute my attribution of a distributed model to Descartes, since resources exist within his account of corporeal memory for dealing with the phenomena in question; and that difficulties in rendering his views consistent result from his uncertainty about exactly what follows from them.

The first objection is this. Descartes cannot have had a distributed model of memory (or any neurophilosophical model), since he believed in an immaterial memory of universals, an intellectual memory more truly human than that of which the mere brain is capable.

Descartes did accept such a memory. But he continually postponed detailed discussion, perhaps well aware of its philosophical limitations.[15] Even if he hoped to integrate theological views on intellectual memory into his overall psychology, his sketchy references to intellectual memory were not taken seriously.

But I take the opportunity to go through four philosophical problems with Descartes' version of an intellectual memory, and in doing so, to develop some positive lines of thought about distributed models of memory. From difficulties about the differences between remembering particulars and universals, we learn something of the relation of exemplars and prototypes in distributed models. From problems about resurrection and the continuity of self, we learn something about personal identity. In correspondence, Descartes resists his own physiological accounts of memory: I look at his difficulties with wonder and with infantile amnesia, and ask why in some contexts he retreated from his speculative neuromechanics. Problems within the intellectual-memory doctrine reveal Descartes' desire for cognitive order, the same desire which sometimes drives him, aware of the fragility of memory, to try to bypass memory altogether.

15 'There are many other points to be noted on this topic [intellectual memory] which cannot now be explained in detail' (to Hyperaspistes, August 1641, AT iii.425, K 112; cf. *Principles of Philosophy* 1.74, AT viii(a).38, CSM 1.221). These postponements are enjoyed by Krell (1990: 61).

Universals and particulars

Burman reminds Descartes that even if, in infancy, 'traces are not imprinted on the brain . . . there still exists an intellectual memory, as is undoubtedly the case with angels or disembodied souls' (conversation with Burman, 16 April 1648, CSM-K 336, AT v.150). Descartes stresses that 'I do not refuse to admit intellectual memory: it does exist' (CSM-K 336, AT v.150). What is the function of this 'altogether spiritual' memory, which 'we mainly use' but which 'is not found in animals' (to Mersenne, 6 August 1640, AT iii.143, CSM-K 151), and what wonders can it perform?

This 'entirely intellectual' memory (to Mersenne, 1 April 1640, AT iii.48, CSM-K 146) preserves only abstract knowledge. It 'has universals rather than particulars as its objects, and so it cannot enable us to recall every single thing we have done' (conversation with Burman, CSM-K 337, AT v.150). Intellectual memory may, for Descartes, be one way to access innate ideas (Foti 1986a: 75; Reiss 1996: 600–1). Since the objects of intellectual memory are 'purely intellectual things, memory in the strict sense is not involved' at all, since 'they are thought of just as readily irrespective of whether it is the first or second time that they come to mind' (to Hyperaspistes, August 1641, AT iii.425, CSM-K 190). But memory is in practice often involved, because abstract things can be 'associated with certain names, in which case, since the latter are corporeal, we do indeed remember them'. This seems to imply that incorporeal memory depends on corporeal memory, but Descartes insists that intellectual memory 'has its own separate impressions', independent of the folds of memory in the brain (to Mersenne, 11 June 1640, AT iii.85, CSM-K 148). Memory of intellectual things 'depends on some other traces which remain in the mind itself. But the latter are of a wholly different kind from the former, and I cannot explain them by any illustration drawn from corporeal things without a great deal of qualification' (to Mesland, 2 May 1644, AT iv.114, CSM-K 233).

The inability of corporeal traces to represent universals was stressed in a different way by Gassendi: while there can be 'a trace of a collection, which by their similarities represent many', this will 'only be an aggregate or composite of many, which have a similarity among themselves'. Corporeal memory can retain only exemplars, aggregates of particulars: something non-physical is required to retain prototypic 'universal natures, e.g. humanity, which are precise and distinct from all grades of singulars' (Gassendi, quoted in Michael and Michael 1989: 43).

The assumption here, shared by Descartes and Gassendi, is that corporeal mechanisms can only connect exemplars by juxtaposition, unable to extract or fuse prototypes out of aggregates of particular cases. How could abstract ideas of triangularity, humanity, goodness, and the like, arise by the mere associative recording of instances, isolated nodes in a memory network? Where would the true forms be in the mind?

But it is exactly the ability to generalise to ideals or prototypes that the Cartesian associative memory was set up to have. The network described in *L'Homme* fills in the rest of a typical face when presented only with a picture of eyes and nose. Automatic generalisation means that the system carries an implicit prototype representation which differs from representations of particular examples (compare McClelland and Rumelhart 1986). Why does Descartes not even address the possibility that his own theory of associative memory might have the resources required? Primarily because the key context to his discussions of intellectual memory is the quite different one of survival and immortality.

Resurrection and personal continuity

In 1642 Descartes comforts the bereaved Huygens with the thought that when, one day, we join the dead in 'a sweeter and more tranquil life than ours . . . we shall still remember the past; for we have, in my view, an intellectual memory which is certainly independent of the body' (to Huygens, 10 October 1642, AT iii.598, CSM-K 216). He protests rather too much the philosophical clarity of this view: the 'very evident natural reasons' for believing it move us more strongly, 'however much we wish to believe', than 'all that religion teaches' (AT iii.599).

One tension in these remarks is pointed out by J. J. MacIntosh (1983; compare Morris 1969: 456). Since all our actions are particulars, and since intellectual memory (which alone survives bodily death) is only of universals, the comfort Huygens should feel in the prospect of eternally remembering the past looks misguided. What is the point of surviving, disembodied, into an afterlife without autobiographical memory of particular experiences in earthly life? What solace will be the recollection of innate ideas, with content which by definition is not tied to the personal history which grief arouses?[16]

MacIntosh's worry can be extended. Eschatology had always been the point of doctrines of intellectual memory. Medieval discussions of the resurrection of the body were consistently materialist, denying that the numerical identity of current and judged individuals could be defended unless their physical parts were reassembled to ensure material continuity: as Caroline Bynum puts it, identity went with the organs (Bynum 1992, 1995b; Davis 1988). But some late scholastics argued, against Aquinas, for a disembodied memory alongside that which depends on corporeal organs, so that the separated soul could recollect the past after death (Kessler 1988: 510; for earlier accounts see Coleman 1992). Separated souls without immaterial memory would be 'crippled for

16 Gordon Baker has suggested to me that Descartes perhaps thought of *conscientia*, thought of as reflection on what is *and has been* in the mind, as providing for the mental memory of particulars which is required here. While this seems a possible view, it is not clear that Descartes accepts or emphasises it: see Baker and Morris 1996: 114.

eternity', noted Pomponazzi (1516/1948: 309) (although his conclusion was not that there is such a memory, but that there are no disembodied souls).

In adding a purely intellectual memory to his physiological model, Descartes thus entered an established theological debate with a respectable conservative line. Officially, for Descartes, I am my incorporeal soul, and so the bodily memory to which I am attached in life cannot contribute to my personal identity. He has no further eschatological concern about responsibility: but exactly this point would lead Locke to use memory as a criterion for continuity of self. On Descartes' view, the fact that I will have only (intellectual) memory of universals in the afterlife does not matter, since I will still be myself, my identity given by identity of soul. To later eyes, this seemed grossly unfair: I may be punished by God for sins (particular actions) I no longer remember committing! The alternative state, in which I might be the puzzled recipient of a prize of eternal bliss awarded on the basis of good deeds of which I have no recollection, came to seem almost as unsatisfactory. In one of the weirder seventeenth-century arguments about memory, Henry More claimed that we *must* in the afterlife retain 'durable traces' of our transactions in this. Only thus could the elect revel in 'peculiar Priviledges of *Morality*' in the heavenly recollection of their pious lives, while enjoying the spectacle of sinners being punished by amoral officers of justice who torture the unrepentant with 'lascivient cruelty . . . according to the multifarious petulancy of their own unaccountable humours' (IS III.II: 187–92).

Problems about the mechanisms of memory came to be increasingly connected with issues of personal continuity. Philosophers would either have to rely on the unwholesomely materialistic orthodox belief in bodily resurrection, or have to find more secure ways to elucidate intellectual memory. With memory thus embroiled in eschatology, difficulties about secular responsibility would also arise: continuity of memory seems necessary to guarantee that the person praised or punished is the same person who committed the criminal or commendable act. In chapters 7 and 9 below I show just how tricky it was to reconcile this theoretical need for continuity with the ubiquitous idea that memory depended on fleeting animal spirits.

Memory and the physiology of wonder

Descartes comes to use intellectual memory to show how an idea of something past is recognised *as* an idea of something past. He writes in 1648:

> If we are to remember something, it is not sufficient that the thing should previously have been before our mind and left some traces in the brain which give occasion for it to occur in our thought again; it is necessary in addition that we should recognize, when it occurs the second time, that this is happening because it has already been perceived by us earlier. (For Arnauld, 29 July 1648, AT v.220, CSM-K 356)

The running of animal spirits down specific impressed paths through the folds of the brain is not, he now claims, sufficient for memory. The mind has to recognise that traces 'left in the brain by preceding thoughts' have 'not always been present in us, but were at some time newly impressed'. Such mental recognition, argues Descartes, proves that 'when these traces were first made it must have made use of pure intellect to notice that the thing which was then presented to it was new and had not been presented before; for there cannot be any corporeal trace of this novelty' (AT v.220, CSM-K 356). So in ordinary perception the mind must recognise ideas as new, as not past.

The import of these passages, beyond Descartes' increasing willingness in certain contexts to argue for 'two different powers of memory' (for Arnauld, 4 June 1648, AT v.192, CSM-K 354), is not immediately clear. Only through a reflection of the intellect at the time of a first impression can we observe a trace to be new (for Arnauld, 29 July 1648, AT v.220, CSM-K 357). But the claim that there can be no corporeal trace of this novelty is contradicted by Descartes' own later discussion of wonder. I know of no other acknowledgement of the connections between intellectual memory and the treatment of wonder in the *Passions*: although I may be missing something which reconciles them, I will briefly sketch the apparent tension.[17]

Wonder, 'the first of all the passions', occurs 'when our first encounter with some object surprises us and we find it novel' (*Passions* II.53, AT xi.373, CSM I.350). There are two causes of the 'sudden surprise of the soul' in wonder: 'an impression in the brain, which represents the object as something unusual', and 'a movement of the spirits, which the impression disposes both to flow with great force to the place in the brain where it is located so as to strengthen and preserve it there' (*Passions* II.70, AT xi.380–1, CSM I.353).

Wonder fixes a local memory trace, as spirits flow between brain, muscle, and sense organs so as to 'continue to maintain the impression in the way in which they formed it'. Wonder is useful 'in that it makes us learn and retain in our memory things of which we were previously ignorant' (*Passions* II.75, AT xi.384, CSM I.354), and in strengthening and prolonging good thoughts which 'otherwise might easily be erased' (II.74, AT xi.383, CSM I.354). The 'novelty' and the strength of the motions of the spirits (II.72, AT xi.382, CSM I.353–4) conspire to isolate a memory trace and render it, temporarily, independent of others. Wonder is a special case in the cognitive economy, uniquely resulting in the formation of a particular place where an impression is located (II.73, AT xi.382, CSM I.354).

In addition to the interest of these passages in the context of the general

17 My reading of Malebranche on wonder and local representation suggested this link: see appendix 2 below. Landormy, Morris, MacIntosh, and Krell say nothing of it. I develop the implications further through a more general discussion of self-control in Sutton 1998.

neurophilosophical model of memory, they imply corporeal traces of novelty. While it can 'perhaps' be through 'an application of our intellect as fixed by our will in a special state of attention and reflection' that the trace of something novel and extraordinary is retained in the memory, 'our idea of it' can also be 'strengthened in our brains by some passion' (*Passions* II.75, AT xi.384, CSM I.355).

In diagnosis of the contradiction, beyond noting that Descartes forgets what the brain can do when wanting to dissociate his theorising from the corporeal, I can only suggest that the deep encoding of a local independent memory trace is, on a distributed model, rare. So, normally, it is difficult for the corporeal mechanism to reveal the novelty of a newly presented object, because new traces are almost always already superposed on a number of other traces in the same fold of the brain. Since a pattern of animal spirit motions through particular brain pores must be reconstructed, it has not always been (explicitly) present, and it is thus hard for the object represented to be recognised as having already been perceived at a particular earlier date. This contrasts with the easy localist account of the perception of duration and of the temporal placement of memories by their location on the coils of memory which critics like Hooke postulated instead (chapter 5 below). Isolated traces may, as the *Passions* suggests, occasionally occur in Cartesian neurophilosophy: but they are too unusual, on a distributed model, to form a basis for a complete corporeal account of recognition.

Infantile amnesia

Infants, claims Descartes, have no 'pure acts of understanding', but only confused sensations, leaving traces in the brain 'which remain there for life', but which are not alone sufficient for memory. 'For that we would have to observe that the sensations which come to us as adults are like those which we had in our mother's womb; and that in turn would require a certain reflective act of the intellect, or intellectual memory, which was not in use in the womb' (for Arnauld, 4 June 1648, AT v.192–3, CSM-K 354–5).

Memory depends on the act of the mind at the time of encoding in addition to its later recognition of the idea as past. Descartes reasserts this: 'if ever I wrote that the thoughts of children leave no traces in their brain, I meant traces sufficient for memory, that is, traces which at the time of their impression are observed by pure intellect to be new'. The case is supported by an analogy very different from *L'Homme*'s analogies of needles in cloth and folds of paper:

> in a similar way we say that there are no human tracks in the sand if we cannot find any impressions shaped like a human foot, though perhaps there may be many unevennesses made by human feet, which can therefore in another sense be called human tracks. (For Arnauld, 29 July 1648, AT v.220, CSM-K 356–7)

It looks as if Descartes, like Erwin Straus, takes this footprints-in-sand analogy to conflict with all notions of corporeal traces: but perhaps it reveals deeper understanding of the nature of distributed traces. Theorists hostile to the trace (because they assume traces must be crudely held in mental storehouses) see the transformations caused by overlaid foot tracks, when there is no exact match between prints and imprinting or trace and past event, as evidence against the existence of traces, rather than as hints at what reconstructed traces might be (Straus 1962/1966: 83–90; Lyons 1981; Krell 1990: 89–91). But in fact this is a good working metaphor for distributed memory. Obliterated but real tracks, those 'unevennesses made by human feet', are the implicit traces, reconstructible if at all only after transformation, providing only indirect access to the past, reconstructing a past which was never present.

Descartes' overt wish in these letters to distance himself from the corporeal-memory model is, indeed, undermined by an earlier idea of his own. He sketched a different approach to infantile amnesia in one of the 1640 letters, noting that 'it is a mistake to believe that we remember best what we did when we were young; for then we did countless things which we no longer remember at all'. But his explanation differs from that of 1648, and is consistent with the theory of corporeal memory.

> Those we do remember are remembered not only because of the impressions we received when we were young, but mainly because we have done the same things again and renewed the impressions by remembering the events from time to time. (To Mersenne, 6 August 1640, AT iii.143, CSM-K 151)

Infantile amnesia is here caused not by the failure of intellectual reflection at the time of experience, but merely through the lack of repetition of the right combinations of spirit motions and patterns of brain configurations. If actions of infants are not repeated, the spirits will not leave the dispositional traces in the pores of the brain which would render recall easier in the future.

Descartes' intellectual-memory doctrine has met with little modern sympathy (Gaukroger 1995: 392; MacIntosh 1983: 346). Defined only negatively, it is parasitic on corporeal memory (Krell 1990: 62). It suffers all the difficulties of general interactionist dualism with conservation of energy and the means of causation across the substantial rift: in addition, I have suggested specific ways in which the official account is undercut by possibilities that distributed corporeal memory could fulfil its functions. I seek not so much another litany of logical problems in dualism, as insight into the particular dangers of corporeal mechanisms which dualists wanted to avert. It could be just because Descartes' psychophysiology gave the self so little autonomy, such a fragile grip on bodily processes, that he sometimes retreated from his own empirical theories.

3.4 Objection 2: memory, soul, and automatism

Memory in machines

How can a dualist have had a neurological theory of memory? Oliver Sacks (1990) has puzzled over the strange fact that a sizeable number of leading neuroscientists believe in a non-physical mind (Eccles 1994). Descartes too saw no tension between seeking imagination and memory in the brains of animals slaughtered by butchers, and arguing that human flexibility in rational action and linguistic response is only possible if we possess an incorporeal soul. Despite insistence in physiological contexts that God the great artificer has so skilfully moulded inconceivably minute textures of matter that the disposition of the organs they compose gives rise to all the behaviour of living animals, Descartes denies that a completed microbiomechanics will explain these key higher functions: we cannot imagine the requisite complexity of mechanism (*Discourse on the Method*, part v, AT vi.58–9, CSM i.140–1). So, it seems, nothing in Descartes' physiology could be a model of human memory, whether reconstructive and distributed or not.

The first answer to this objection is that Descartes does not take memory to be a *mental* capacity at all, for it does not require consciousness or soul. Its explanation is embedded among accounts of life functions: God could (and for us has) joined in puzzling mixture a rational soul to *L'Homme*'s living body-machines, but their life processes operate independently of that union. Descartes is clear:

> the effect of Memory which seems to me to be most worthy of consideration here is that, without there being any soul present in this machine, it can naturally be disposed to imitate all the movements that true men – or many other similar machines – will make in its presence. (AT xi.185)

This is not 'memory' in any derivative sense: this just is (corporeal) memory. Only intellectual memory is unique to humans: the celebrated beast-machine doctrine (*Discourse*, part v; to Plempius for Fromondus, 3 October 1637, AT i.413, CSM-K 61–3; to Newcastle, 23 November 1646, AT iv.573–6, CSM-K 302–4; to More, 5 February 1649, AT v.275–9, CSM-K 365–6) does not deny corporeal memory to animals. Soul-less machines or beasts have many capacities which *we* would class as cognitive. Even though Descartes does not so class them, the various phenomena he tries to explain are genuine phenomena of memory, sensation, emotion, imagination, and so on. These are the functions which the strangely powerful spirits and nerves can perform. Knowledge of the physiology here, Descartes thinks, in fact shows us how little our soul is responsible for. So although there can be no Cartesian science of mind, there can and must be a science of memory.

But this answer is unlikely to satisfy. Surely, without a soul, the only 'memory' possible would be reflex action, mere automatism, not the personal

memory we care about. Is the fact that Descartes' physiology of memory excludes consciousness not enough for us to dismiss it as unsatisfactory: is it not 'clear that the one thing Descartes was *not* explaining was the *psychological*-ity' of what we call psychological events (Richards 1992: 65–6)? It is just because Descartes sees animal action as inflexible, wholly stimulus-driven, that he thinks beasts are automata (Seris 1993). On this view, humans have two modes of response to the world, automatic reflex behaviour, imposed by passive brains soaking up stimuli, and incorporeally-mediated (because conscious) action, while beasts have only the former.

I address the objection in this form. It assumes falsely that the dichotomy between reflex action and true *action* derived ultimately from the soul is exhaustive, that you are either an automaton or a rational soul. In fact, Descartes accepts an intermediate form of interaction with the environment, a class of responses into which fall many phenomena of great interest to him and to us.[18] To defend this controversial claim, I start with an example of the orthodox interpretation which I dispute.

Reflex action, automata, and beast-machines

Owen Flanagan (1991: 3) describes the impoverished world of the 'Cartesian automaton', restricted, because it is only body, to automatic reflex behaviour in dealing with the world:

> the complete system of wired-in reflex arcs exhausts its behavioral potential. What a particular automaton eventually does, how it in fact behaves, is the inevitable result of the interaction between the environment and the wired-in arcs. Such a system is deterministic in the sense that, barring mechanical failure, there is one and only one response for each stimulus.

This is at best misleading, more probably false. Flanagan implies that automata will always respond in the same way to the same stimulus, and indeed that all (similarly designed) automata will respond likewise.[19] Only initial wiring,

18 A related response to the objection, which I do not pursue here, is to accept that Descartes' corporeal memory includes only the forms of memory which psychologists now call 'implicit', with no conscious awareness of remembering (Jacoby 1988). Even if Descartes excludes consciousness (whatever that is) from bodies, implicit memory is vital not only for successful ongoing life processes but also as a central source of explicit subjective remembering in beings like us who also have minds. The effects of implicit memory can feel like 'mental contamination' as we are often unaware of the sources of our remembering (Schacter 1996: 161–91). Descartes' theories of implicit memory are intended to help us in the troublesome task of extending control into the systems which underpin memory-without-awareness, to exchange hard conscious intrusions into associative processes for the common, haunting, intrusive effects of implicit memories on consciousness.

19 Flanagan continues: 'or to put it another way: given any stimulus, it is impossible for the system to do other than it in fact does.' This is a red herring. Descartes probably thought it true *even for rational human action* that 'given any stimulus, it is impossible for the system to do other than it in fact does.' At least in the Latin *Meditations* (AT vii.58), he denies that human freedom entails the positive 'two-way' power of determining to do either of two

fixed before the automaton goes out into the world, and immediate present environmental input count: so of course all automata will always react in the same way. The behavioural potential of *some* simple organisms is indeed exhausted by its system of wired-in reflex arcs, with no learning from experience or modification of wiring in new environments. But this is not the case for Cartesian automata (and it is certainly not implied in the very idea of an automaton).

Flanagan, in fact, sets up the issue wrongly by looking at Descartes' philosophy of the body through an 'analogy' Descartes allegedly draws between inorganic automata and the organic human body (1991: 3). The problem is not just that everything artificial is, for Descartes, *also* natural (*Principles of Philosophy* IV.203, AT viii(a).326, CSM I.288): it is that automata, specifically, *are* organic, self-moving machines which behave as they do because of the disposition of their *organs*. This is not to eliminate the organic (or the cognitive), but to explain it by reduction (Gaukroger 1995: 279, 287–90). The organic nature of living human bodies does not make them, for Descartes, any less automatic (compare Carter 1985/1991: 207–8). Yet the 'disposition' of their organs is not fixed at birth.

The case of memory makes this clear, for in memory processes there is no immediate motor response to stimuli, and if there is a delayed motor output it is not hard-wired (Gaukroger 1995: 280–1). Corporeal memory transmits effects of experience over long temporal gaps, so that they are causally involved in behaviour mediated by complex internal processes. As Descartes notes, it is not necessary for the incorporeal soul always to be implicated in these processes. Yet the determinism involved in these responses is not a simple stimulus/response link, for corporeal causes act holistically, and are not restricted to those immediately current. To put it another way, memory shows that an automaton's physiology changes over time. Automata with different histories, different 'experiences' marking their brains and bodies, will respond differently, and one automaton will respond differently at different times to the same stimulus after new experience has modified the spirits and pores in the folds of its brain.

Conceptual room for this intermediate category between the simply automatic and the incorporeally free can be found at a number of levels in Descartes' texts.[20] I approach it through the neurophilosophy: a schematic outline will help.

contraries, though Michelle Beyssade (1994) suggests that he did accept this incompatibilist Jesuit view of freedom as 'indifference' after 1645 (compare Chappell 1994). So of course Descartes denied free will to automata: but this is not at all equivalent to Flanagan's much stronger claim that an automaton's behaviour is wholly determined *only* by wired-in reflex arcs and current stimulus. Determinism allows greater complexity than that.

20 In the distinctions between three grades of sensory act (replies to the sixth set of objections, AT vii.438, CSM II.295–6), the first grade is physical and physiological and purely automatic; the second is sensory, in some way the immediate mental result of the mind's union with sense organs; and the third is the grade of judgement. But these distinctions are far from clear, and the intermediate mode of cognitive response which I

Organism/environment interaction: three possible forms

	Hard-wired	Soul	Pineal
1. Simple automatism	Yes	No	No
2. Complex automatism	No	No	Yes
3. True action	No	Yes	Yes

Simple automatism, firstly, is hard-wired, an immediate and invariant response to a sufficient stimulus. Even if Descartes' account of the reflex is 'short on detail about the specifics of neuroplumbing' (Hatfield 1992: 348), the following seems clear (Canguilhem 1955/1977: 27–56; Spillane 1981: 84–9; Clarke 1989: 185; Dennett 1991a: 105, 321–2). Automatic behaviour is mediated, after stimulus information is transmitted from the sense organs, by animal spirits flowing from the brain ventricles into the muscles. It is 'just the same' as the immediate, fixed chains between the passage of air through organ pipes and the particular sounds the organ produces (AT xi.165–6), and this analogy proves that Flanagan's account is appropriate here: 'the organ simply produces the music as a result of an input: it does not represent the notes to itself, in the way that the organist might' (Gaukroger 1995: 280).

Simple automatism does *not* involve the pineal gland: the switch from sensory to motor response occurs when the entrance to a brain pore or tube is opened by the motion of a nerve fibre, and animal spirits from the ventricles enter and are carried through the tube to various muscles (*L'Homme* AT xi.142, H 34–5, CSM i.101–2). Some commentators have been confused into attributing to the pineal gland a role in reflex action by the much reproduced illustration of an automatic response to fire near the foot (Figure 3).[21] But others (such as Jefferson 1949: 699) have understood that the oval form in the brain labelled F is described clearly in the text as 'the cavity F', and thus, in Descartes' vocabulary, can only be the brain ventricle and *not* the pineal gland (Figure 4).

So memory ideas are traced in spirits on the surface of the gland: but the gland is not involved in simple automatism. The difference between reflexes

attribute to Descartes would include both the second grade of sense and some cases of the third, those in which judgements are made on the basis of habits formed in childhood in response to the institutions of nature (rather than the mature judgements of the rational mind). Reed (1982: 736–7) misleads by limiting the third grade to what is 'essentially mental' and not allowing for this division within it. For a better interpretation see Hatfield 1986, 1990: 44–5. All I rely on here is a wider class of responses which do not involve the soul, but which are also not simple automatism.

21 More emphasis can be placed on the textual reference to a cavity than on these pictures, which were not Descartes' (AT xi.vi–vii). But the pineal gland is consistently labelled H (*not* F). Descartes' later discussion of the different neural effects of placing a hand near moderate and hot fire (AT xi.191, H 102–3) refers only to the differential opening of tubes in the brain mesh and on the internal surface of the ventricular cavities, and *not* to the gland.

Figure 3 Familiar image of reflex action from *L'Homme* (Descartes 1664: 27), which has encouraged the mistaken identification of 'the cavity F' with the pineal gland

and corporeal memory is that reflex pathways are unalterable, whereas the passage of spirits through the pineal gland drives them into regions of the brain which do change, allowing the plasticity of memory. Therefore there are no ideas or representations, strictly speaking, in simple automatism. This is sensation and immediate response without representation: a primary function of representations, to extend capacity for response over time, is not required.

Actions caused by the soul,[22] at the other extreme, *do* require the mediation of the pineal gland to transmit or direct its action in the physical world. But, and

22 The language of causation here (from mental to physical) seems unchallengeable even by occasionalist interpreters of body to mind relations. John Yolton, who prefers to think of Descartes' physiological events occasioning (rather than causing) sensations, points out (1990a: 72, n. 30) that non-causal occasional or significatory relations apply only in the direction of body to mind, since 'the mind seems able to cause or make the pineal gland move and hence affect the body's physiology'.

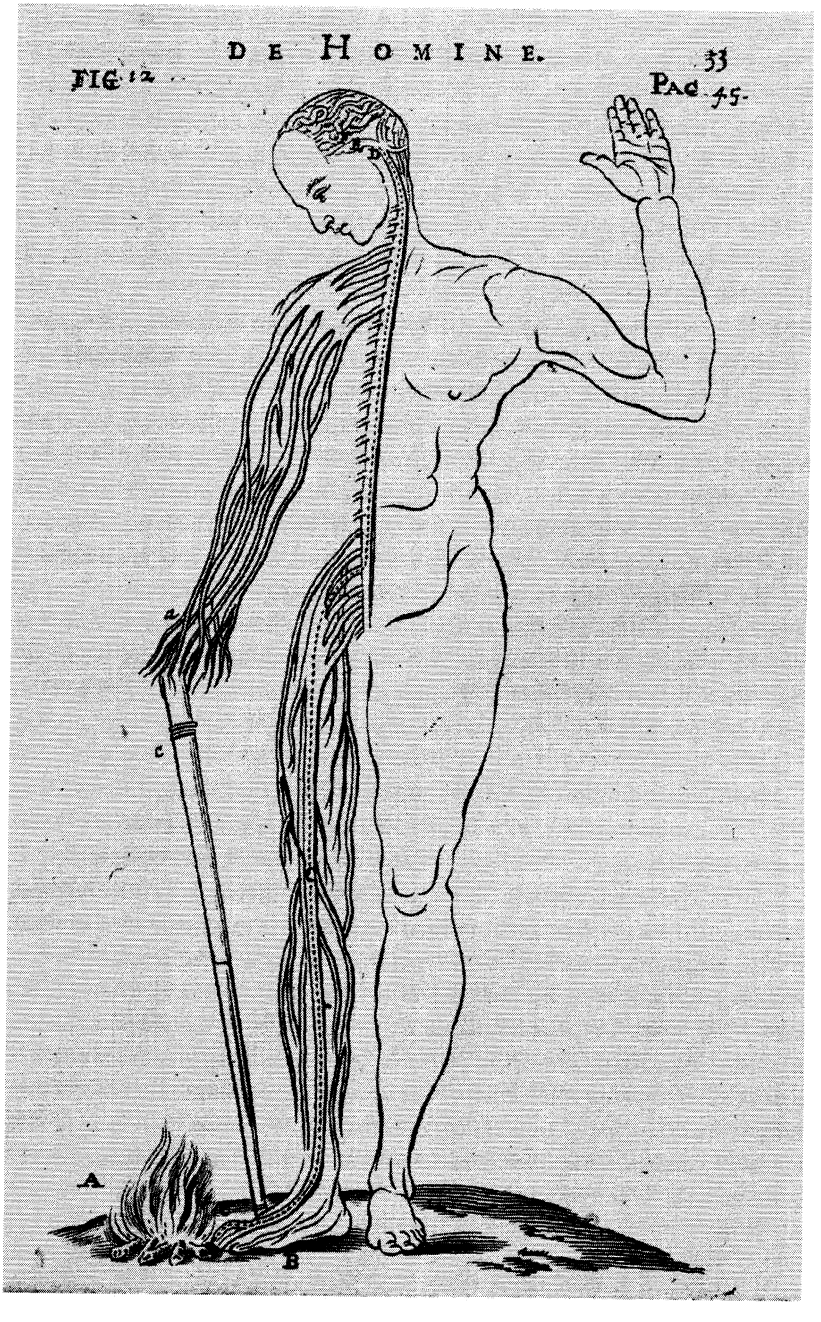

Figure 4 Simple automatism: this strange illustration of a Cartesian automaton, from Schuyl's Latin edition of *L'Homme* (Descartes 1662: 33), makes it easier to see that 'the cavity F' is the ventricle (Wellcome Institute Library, London)

this is the vital point, there are also cases of behaviour in the causation of which the pineal gland is involved where the soul is not.[23] Thus there are three separate kinds of response. As well as (1) simple automatism and (3) true action, there are intermediate responses (available to organic automata) which could be called (2) complex automatism, 'physical-cognitive', or even 'cognitive-automatic'. Remembering is not *simply* automatic: this is *both* because it involves the pineal gland, and because it can occur over long interrupted periods, longer than the longest kinds of stimulus/response arc. Again, it need not involve the rational soul.

There is extra evidence for this threefold schema in Descartes' attitude to explanation by reflex. He is aware of hardwired immediate unconditioned reflexes in humans and animals, when sheep run from wolves, or when humans throw out their arms when falling 'without the assistance of any soul' (replies to Fourth Objections, AT vii.204, CSM II.144). But there are also much longer term, yet still wholly physical, responses in which corporeal memory is at work. Some are cases which we would call 'conditioned response', the acquisition of learned associations where there is no natural 'relation' between a representation and its 'meaning'. 'If you whipped a dog five or six times to the sound of a violin, it would begin to howl and run away as soon as it heard that music again' (to Mersenne, 18 March 1630, AT i.134, CSM-K 20). Setters can be trained, against natural inclinations, to stop at the sight of a partridge and run towards it on hearing a gun (AT xi.370, CSM I.348).

But rather than conceptually isolating these conditioned responses of dog-machines or linking them with simple automatism, Descartes couples them, both in 1630 and in the late 1640s, with more complex human cases which he considers equivalent. The dog howling at the music of whipping is linked with an account of why 'what makes one man want to dance may make another want to cry', when the latter has 'never heard a galliard without some affliction befalling him': he cries 'because it evokes ideas in our memory' (AT i.133–4, CSM-K 20). The case of the trained setter, in the crucial sections of the *Passions* on psychological conflict and self-control, is 'worth noting in order to encourage each one of us to make a point of controlling our passions. For since we are able, with a little effort, to change the movements of the brain in animals devoid of reason, it is evident that we can do so still more effectively in the case of men' (*Passions* I.50, AT xi.370, CSM I.348).

23 The point of the pineal, then, is not (as caricatures of Cartesianism would have it) simply to buttress metaphysical dualism: animal automata have pineal glands too. Jaynes' claim (1970: 226–7, followed by Harrington 1987: 6) that Descartes thought that animals had no pineal gland is false: Descartes was accustomed to finding it 'without any difficulty in freshly killed animals'. He had more trouble locating the metaphysically crucial gland in human autopsies (to Mersenne, 1 April 1640, AT iii.49, CSM-K 146).

Not everything, then, in humans or in dog-machines, is innately wired in, for 'the movements of the brain' change in the course of experience. Cartesian automata are not the uncanny 'Neurospasts', nimble sprightly puppets which only *seem* to be moved from within, feared by vehement defenders of free will like Cudworth and More (Gabbey 1992: 117). It is just because wholly corporeal long-term conditioning and the long-term workings of associative memory (which Descartes classes together) are flexible in these ways that knowledge of them is so important to those unique beings, ourselves, who do also have souls. The soul's occasional influence on bodily states is its only way of instituting better habits in the wayward dynamics of spirits and brain. Associative memory is the key capacity for us to use in the moral attempt to divert our own internal fluids from running into dangerous traces: we can *learn* 'to separate within ourselves the movements of the blood and spirits from the thoughts to which they are usually joined' (*Passions* III. 211, AT xi.486, CSM 1.403; see also Barnouw 1992: 406–8).

The explananda in *L'Homme*'s soul-less world extend well beyond immediate reactions to stimuli, covering intervening variables like the long-term interactions of corporeal memory traces. There is no incompatibility between all the complexity of the life functions and mechanism (Mackenzie 1989: 168): but the life functions extend well beyond innate reflex arcs. In machines with the right microstructures, the past leaves specific traces in changing motions inside the body. Only physical variables are involved, and yet the notions of experience and individual learning history are applicable. The diverse causal factors involved in registering and integrating information include 'previous brain episodes' and non-neural bodily events as well as current environmental input: 'this is the model of an automaton, to be sure, but not one which operates by reflex' (Mackenzie 1989: 174–5; compare Fearing 1929: 386–7; and Grosholz 1991: 126–7).

This is exactly right: there are response which are 'cognitive' or at least not simple-automatic, which involve the rolling pineal gland, but which do not require the soul. This is enough to correct some unsympathetic criticisms. Marjorie Grene (1985: 47–8) is indignant that Descartes ascribes memory to animals after so foolishly claiming that beasts are machines. But when a conceptual space *between* the automatic and the non-physical is opened, it is no surprise that impressions can be formed in the brains of animals by, among other things, 'the traces of previous impressions left in the memory, or by the agitation of the spirits which come from the heart' (to Princess Elizabeth, 6 October 1645, AT iv.310, CSM-K 270). Generally, contrary to popular views (Williams 1978: 282), Cartesian beast-machines and living automata *can* feel, remember, imagine, sense, and dream (Rodis-Lewis 1978; Gaukroger 1993b, 1995: 287–9, 392–4; Baker and Morris 1996: 91–100).

3.5 Objection 3: mechanism, matter, and active bodies

Inert bodies? In their call for papers, the organisers of an interdisciplinary philosophy conference on 'Body Matters' complained (Burwood and Jagger 1994) that 'The Cartesian legacy has furnished contemporary thinking with a paradigm of the body as an inert, closed, and anonymous object.' A curious consensus in analytic history of philosophy, medical anthropology, feminist theory, and cultural studies at large coalesces around the image of Descartes as anti-magus, stripping nature and the human body of all powers and activity. An earlier enchanted world, criss-crossed by networks of sympathies and antipathies, embracing analogy and suggestion over representation and intervention, traversed by holistic herbalists and natural magicians, coupling earthy bodily realism with organicist ecologism, was sundered and lost with Descartes' blind scientistic drive for the mastery, possession, and penetration of nature. The removal of mind from body is but one symptom, on this view, of multiple damaging Cartesian schisms, dividing philosophy from biology, science from history, power-mongering manipulators of nature from the dead ecology which they exploit, active rational male observers from passive fragmented female bodies, or mechanistic interventionist medicine from a more watchful psychosomatic physic of the whole person (e.g. Toulmin 1990: 107–15).

From this line of thought springs the third objection to my interpretation of Descartes on memory. Distributed memories are always in motion, never faithful in inert cells in a memory palace. Surely, then, the Descartes we all know and love to hate *cannot* have had a dynamic distributed model of memory: his mechanistic natural philosophy notoriously renders the body and its contents, like all extension, purely passive, moved only by the active power of mind or other disembodied intelligence. Specifically, Descartes notoriously reduces (or wishes to reduce) the whole of natural philosophy to mechanics, allowing no dynamics and no active forces: how then could he imagine memory motions superimposed and re-separated in the intuitively non-linear fashion characteristic of distributed models?

I respond by denying the existence of such a clear line in Cartesian natural philosophy between the passive and the active. Despite rhetoric about inert matter, mechanism was (and is) compatible with activity in physical and biological worlds. Immersion in Descartes' neurophilosophy shows that, despite twentieth-century historiography, more 'dynamic' concepts of disposition and pattern have unproblematic mechanistic senses. The human body is a multiply constrained composite, and is not in practice inert: general Cartesian physiology is a dynamics of fluids, a brave attempt to listen to the pulse of the body-machine.

These are large claims, which I overstate to correct easy assumptions of their falsity. I sketch the form of their defence in three directions. I describe the

rationale for thinking of Cartesian natural philosophy, and in particular physiology, as (a form of) dynamics. Then I return to the question of machines, arguing that seventeenth-century mechanism was not, as Carolyn Merchant, Otto Mayr, and others in diverse fields claim, intrinsically authoritarian: Cartesian automata are not wholly reliant on external design and direction, but have their own activity, specificity, and causal powers; they lack only the acausal autonomy and authoritarian control over body and brain officially attributed to the soul. Finally, I generalise this point to describe the puzzlingly anti-'modern' nature of the human body in Descartes' neurophilosophy: mechanisation does not eliminate but exacerbates the permeability of the open body.

My motivation here is not to deny the continuing political and intellectual need to undermine the simplifying dichotomies I listed above. They have had real and often damaging effects, structuring thought, feeling, and action, though in many cases later than the seventeenth century. Again, my caricature of fetishised 'false nostalgia' for 'some lost, but recoverable, perfection' in the pre-Cartesian world (compare Reiss 1996: 592–3) is not intended disrespectfully, for the very romance of a pre-modern golden age is the most powerful initial encouragement to seek glimpses of more complex historical differences. But I seek more nuanced understanding both of the target positions and of the space of possible alternatives, through historical dismantling or dissolution of easy stories about their 'origins'. This is not only misplaced scholarly rehabilitation of a maligned metaphysician for a world heavier with grief, anger, and decay than was even his. It is also driven by a desire to test the possibilities and limits of current flirtations with the fragility of memory, by questioning our common visions of the earlier ideas of science and the body which supported and clashed with the image of fluid motions in the folds of memory.

3.5.1 *Dynamics, mechanism, and Cartesian physiology*

Descartes, judges a historian of physiology, was 'a representative of the baroque, partial to a dynamic interpretation of nature' (Rothschuh 1953/1973: 78; compare Jefferson 1949: 692). This sounds odd to philosophers and historians of science reared on sharp contrasts between the barren world objectified by Cartesian science and the more sympathetic, nested worlds of baroque Leibnizian monads (Deleuze 1993) or Newtonian forces, ethers, and active powers.

Descartes did pursue and extend the anti-secularising project of Mersenne and other Paris mechanists who sought to drain nature of the forces and sympathies of Renaissance Neoplatonists and naturalists, combating above all the atheism implicit in attributing too much autonomy to matter: mechanism was initially attractive, in the 1620s, because of its supernaturalising theological conservatism (Merchant 1980: 196–200; Hutchison 1983; Hine 1984;

Gaukroger 1995: 146–52). But other forms of activity, differently justified and explained, remained in the mechanical cosmos. It is not quite true that in Descartes' work 'all spirits were effectively removed from nature' (Merchant 1980: 204). The survival of paradoxically corporeal animal spirits was not an accidental residue, a pun uneasily transmitted between organicist and mechanistic worlds. Their incessant motion is genuine activity, whether or not it ultimately derives from God: their coalescences, breachings, foldings, and commotions, retaining and transforming patterns over time, can continue quite independent of the individual will.

Historians of science have recently demonstrated the intrinsic roles of activity, sympathy, spirit, and force in later seventeenth-century English 'mechanism' (Henry 1986a; Schaffer 1987; compare Hutchison 1991), and I extend this research in discussing problems of action at a distance in chapter 4. But if pre-Newtonian English mechanism was not just compatible with activity in nature but required it, I can apply a similar strategy in Cartesian historiography, where there is serious difficulty in understanding quite what 'strict mechanism' could have been. Descartes' explanatory ambition did not require the *elimination* of puzzling and complex natural phenomena, as implied by Merchant (1980: 204–5): with strange facts of the organicist world, he sought reduction, *not* elimination. It was not, in many cases, the baffling phenomena (the bleeding of wounds on the approach of the murderer, the weapon salve, sympathies, the maternal imagination imprinting on the foetus) which he rejected, but only certain candidate explanations of them, such as those which attribute thought or free will to corpuscles (*Principles* IV.187, AT viii(a).314, CSM I.278–9; Shea 1991: 111–20). These are traditional examples of the activity of matter, and Descartes denies not that they occur, only that it is necessary for matter to have the extremely strong capacity of free rational thought for it to exhibit such behaviour. This is true too for human science: reason, thought, and freedom are unique to immaterial souls, but they are not by any means everything. Passion, dreams, confusion, and long-term reconstructive memory were among the explananda left over for a genuinely dynamic physiology of mind.

Real forces But, in a difficult historiographical debate, some scholars of Cartesian physics claim that Descartes had no dynamics, because he did not admit forces in natural philosophy: force in Descartes' physics is just God, sustaining across discontinuous instants corporeal matter in motion, which has no powers of its own, according to a principle of continuous creation (Hatfield 1979; compare Westfall 1971: 56–98; Gabbey 1980; Gueroult 1980). It is far from clear what this metaphysical doctrine entails in natural philosophy. Stephen Gaukroger argues forcefully for a minimal reading of Descartes' official commitment to inertness, as requiring only 'that matter does not *initiate* any kind of activity': this allows that forces are *sited* in matter although

ultimately *derived* from God (Gaukroger 1995: 376–7, my emphasis). In contrast to the popular view (based on the atypically strong claims of the Third Meditation) that Descartes eliminated forces from nature and sought to reduce physics to kinematics (Williams 1978: 261–2), Gaukroger believes that 'force is built into his account at the most fundamental level' as real tendencies to motion (1995: 247, also 70, 83–4, 343–4, 375). Readings which subordinate Descartes' view that bodies have *within* themselves the force to continue to move or to remain at rest (to Mersenne, 28 October 1640, AT iii.213, CSM-K 155) to impose a 'consistent' metaphysics of instantaneous divine action neglect the developmental sequence of Descartes' thought (physics before metaphysics), and ignore his advice to pursue desirable 'physical studies' rather than elaborate metaphysical questions (Conversation with Burman, 16 April 1648, CSM-K 346–7, AT v.165; compare to Princess Elizabeth, 28 June 1643, AT iii.691–5, CSM-K 226–8; see Gaukroger 1995: 10–14, 375).

Many disagree with Gaukroger's claims for forces among the corpuscles. But, just as he cites centrifugal forces in Cartesian cosmology as evidence that 'one could only believe God was the sole site of activity if one concentrated exclusively on Descartes' metaphysics' (1995: 375–6), so my discussion of physiology suggests wider implications of the case for real Cartesian forces. I address Emily Grosholz's critique of *L'Homme*, as a specimen in physiology of the common reading which takes the subordination of physics to metaphysics to be *Descartes'* problem rather than the interpreter's. It is worth touching first on two related (and less controversial) issues in the general natural philosophy: *hydrodynamics* and the *plenum*. Both are given new emphasis in Gaukroger's work, and both support the dynamically tinged Cartesian physiology.

The fluid-filled cosmos While working with the microcorpuscularian theorist, Beeckman, in 1619, Descartes wrote about the behaviour of fluids. His manuscript furnished 'certain concepts and modes of argument . . . which will constitute the essence of Cartesian micro-mechanism in optics, cosmology, physiology, and natural philosophy generally' (Schuster 1977: vol. 1, 94, quoted by Gaukroger 1995: 84; AT x.67–74; Shea 1991: 27–33). The mechanism of the early 1630s remained 'very dependent upon a hydrostatic/hydrodynamic model' (Gaukroger 1995: 225). Solid and fluid bodies are not metaphysically distinct, but on a continuous spectrum, with solids as conglomerations of corpuscles which are closer to rest with respect to one another (*Le Monde*, AT xi.11–15, CSM 1.84–5). Descartes illustrates his claim that 'all the motions that occur in the world are in some way circular' with the example of a fish in water, where motion is by displacement (AT xi.19–20). Cartesian astronomy rests on a 'hydrodynamic model of the cosmos, which requires solid bodies, such as planets, to be embedded in a fluid which carries them along in a vortical motion' (Gaukroger 1995: 412, 249–56; AT xi.50–83; compare Rodis-Lewis

1978: 160 on dynamics in mechanism). Descartes' account of planetary motions by the whirling mutual displacements of contiguous vortices subsumes cosmology into hydrodynamics, 'and hence his interest is really in fluids' (Gaukroger 1995: 234, see also p. 247).[24]

The fluid model, this physics of circulation, displacement, and endless motion, is secured by rejecting the void in favour of a plenum (Hesse 1961/1970: 102–8; Heilbron 1982: 22–6). It is natural to think that mechanistic reductionism, whereby all observable bodies are composites of small particles *must* go (as it did, for example, in Gassendi) with the ancient atomists' ontology of isolable atoms scooting or swerving, alone or in packs, across a void. But this is not so. In 1629 Descartes expressed agreement 'on the whole foundations of Philosophy' with Sebastian Basso, a critic of Aristotle who combined atomism with a neo-Stoic plenum theory (to Mersenne, 8 October 1629, AT i.25).[25] In *Le Monde*, moving bodies are always surrounded by other bodies. So everything affects everything else, indirectly if not by one of the direct collisions which are the constant fate (or chance) of every conglomeration of material elements (Gaukroger 1995: 230–1, 275).

As Gaukroger points out (1995: 241), it is easy to imagine here the intuitive Lucretian vision of atoms or composites moving for long stretches through a void and only occasionally colliding. But this is where the fish example breaks down, for there is no fluid-free part of the plenum. Cartesian fish are never out of water, since bodies are always colliding and are never out of mutual causal contact. The combination of hydrodynamics and plenum physics has intense intuitive power, in two related respects which I use to bridge the gap between physics and physiology: *context-dependence*, and *causal holism*.

Context and complexity In the cartoon version of mechanism which I sketched at the start of this section, the machine metaphor lies behind a deadening inattention to emergent behaviour, to the way wholes act differently from their parts. Mechanism as method is the analysis of complex wholes into

24 The tempting thought that this talk of the metaphysical identity of solids and fluids (compare *Principles* II.54, AT viii(a).71, CSM I.245, *Description of the Human Body* Part III, AT xi.247, CSM I.319) is in fact a blanket reduction of fluid to solid, to reabsorb fluidity in particulate form, finds support in quite different historiographic traditions (Shapiro 1974; Irigaray 1974/1985b). But I suggest that, in the thoroughly permeable Cartesian cosmos, the diffuse, boundary-blurring, temporally bound fluids seep across all Descartes' natural philosophy.

25 Basso thought that any possible spaces between the smallest particles were filled by a 'universal spirit', something like the Stoic *pneuma*. This was Descartes' one point of difference: 'I do not explain the ether as he does' (quoted in Gaukroger 1995: 221). His form of 'cosmobiology' did not require the *identity* of physiological spirits and cosmological matter. Barker and Goldstein (1984) take Descartes' physics as proof that atomism and the plenum were not exclusive options, and trace the transmission of neo-Stoic physics up to and including Basso and Descartes (see also Barker 1991 on the mechanising of Stoic *pneuma*).

discrete parts, treated in isolation. Critics of mechanism then complain that the nature of bodies, and of events, is correspondingly taken in totalising fashion to be independent of the particular context (physical, social, or temporal) in which they exist or occur (Merchant 1980: 227–35). This context-independence is indeed, crudely, the ideal fiction of atomist kinematics, in which isolated collisions in the void between bits bashing each other model more complex interactions within or between other bodies. But, Gaukroger argues (1995: 241), Descartes' approach is not kinematic but hydrostatic: 'the point seems to be not so much to analyse the behaviour of a body under various kinds of constraint in terms of how it behaves when not under constraint, but rather to account for what happens when a body moves from one system of constraints to another'. In 'body matters', then, Cartesian bodies (human and other) are not isolated or closed: Descartes shows 'almost no interest in unconstrained bodies', but assumes that 'systems of constraint are constitutive' of the phenomena under investigation (Gaukroger 1995: 247–8).[26]

The distributed model of memory sketched above is firmly embedded in this natural philosophy, for in memory too all motions and events are wholly context-dependent. As patterns rise in succession in the folds of memory, their particular cast depends not only on current input from world, body, or soul but also on the contingent dispositional states of the pores in those regions of the brain, and on all the messy factors influencing the state of the animal spirits. Cartesian memories share not only this context-dependence with the physics, but also a second property, that of causal holism. Because different figures are traced implicitly in the same folds of the brain, and because animal spirits can only reconstruct patterns in the present, there is a clear sense in which all (or at least many) traces are causally active in the (re)construction of any one pattern during any memory event.

This is just a special case of the causal holism which, like context-dependence, simply falls out of Descartes' natural philosophy. In the full fluid

26 Merchant (1980: 230–1) quotes *Principles* II.53 (AT viii(a).70, CSM 1.245), where Descartes says that it is hard to apply his laws of motion 'because each body is simultaneously in contact with many others', and that 'no bodies in the universe can be so isolated from all others'. But this is precisely Descartes' reminder (in a more defensive statement of the laws than that of *Le Monde*) to think of collisions as continuous in a plenum rather than occasional in a void: he is not, as Merchant suggests, unwillingly 'forced to admit' this context-dependence. Jamie Kassler (1997) argues strongly for a sharp contrast between Descartes' and Hobbes' plenum theories, and favours Hobbes' (truly neo-Stoic) picture of tensional forces filling space, including the invisible internal motions of physiological 'endeavour', over Cartesian atomism in which all change of motion in bodies is from the outside by impact and pressure. But Descartes is closer to Kassler's interpretation of Hobbes than she allows. It is not just that 'impact' and 'collision' are inadequate descriptions of the continual circulation of Cartesian bodies in the plenum, but that the example of memory in Cartesian automata proves that changes of motion arise internally, with physiological tendencies to motion maintained over long periods without specific external intervention.

cosmos, any motion anywhere is also necessarily motion elsewhere, for no body is unconnected or unconstrained. As for the Stoics, causation is conserved throughout the integrated cosmic system, with everything holistically linked to everything else (Gaukroger 1995: 242–3). Memory motions, like all material interactions, are embedded in multiple conspiracies of causes.

Real patterns Psychophysiological dynamics are of a piece with the complex interconnected meshings of forces throughout the Cartesian universe. But critics claim that the dynamic elements which causal holism imports into Descartes' physiology are somehow illegitimate, 'smuggled goods' or tacit annexations which enrich with forces and thus violate a basically pure official static mechanism (Rodis-Lewis 1978: 152–4). In a well-constructed critique which opens up a set of difficult issues, Emily Grosholz (1991: 118–19, 127–8) complains at Descartes' invocation of non-linear feedback systems which are hardly analogous to the simple impact machines on which his physics was meant to be based. She is, perhaps, too carping in blaming Descartes for theorising non-linear systems which are not quantifiable 'by means of the mathematics he knows' (1991: 128). When, as I do, one has independent motivations for interest in deterministic but non-linear dynamic models of memory, there is ample cause to seek keenly their perceived implications.

Grosholz follows, with modifications, the tack of interpreting all Cartesian texts by the methodological standards of the *Meditations*. Despite Descartes' point that theories in natural philosophy are not uniquely constrained by metaphysical principles (Hatfield 1985), she attributes to him the desire always to move in physics back from complex bodies and phenomena to the sparse set of simple natures and elements justified in metaphysical reflection. His 'unexamined faith' in reductive analyses, argues Grosholz, leaves Descartes unable to explain complex entities and processes, imposing unnatural homogeneity on diverse phenomena, impoverishing domains in which the action of complex unified wholes is irreducible to the dull mechanical interaction of their isolated parts.

Physiological and cognitive phenomena are subject to one unnatural reduction of natural complexity imposed by this Cartesian 'method at odds with itself' (Grosholz 1991: 118). Not only is Descartes' use of 'animal spirit' suspiciously unmechanistic in spiritualising subtle matter (1991: 122): he further invokes as analogies, unfairly, not simple machines but feedback mechanisms to suggest that cognition, like respiration, is 'really a highly sophisticated feedback mechanism' (1991: 126). This applies notably to the idea of corporeal memory as diverse or patterned motions, and to Descartes' use of the pineal gland as 'a locus for complex algorithms which convert perceptual patterns into patterns inducing movement' (1991: 126–7).

This key concept of the 'pattern' is essential for the neurophilosophy to get

off the ground. 'Pattern' and the related term 'disposition' are something like second-order physical properties, properties which supervene on first-order physical properties like size, shape, rotation, and so on.[27] But Grosholz argues that, given his officially meagre mechanical ontology, there is in fact 'no room ... for the notion of pattern which is so central to Descartes' cognitive physiology' (1991: 127). The need for 'more or less permanent configurations in the fibres' to ground his theory of memory forces Descartes to 'either revise or short-circuit his original conception of causal interaction' (1991: 126–7).

The problem here is not only that Grosholz overemphasises the permanence of patterns in the spirits moving through folds of memory, nor that she just rejects the reality of forces sited in matter, though I have suggested reasons for differing on these points. More specifically, I think that Grosholz's case *assumes* the central point at issue: she simply *denies* that animal spirits which are merely bits of matter in motion could *also* be 'active and patterned, conveying information' (1991: 129). But the contention of the cognitive scientist (Cartesian or connectionist) is that items which interact *only* in virtue of their physical properties *can* also maintain *some* regularity of pattern through the course of such interactions. Patterns do *not* have to be transmitted '*qua* patterns' in order to have causal effects (as Grosholz says they would), because they can be transmitted by ordinary physical-causal means.

Grosholz makes the strong contrasting claim that 'for a pattern to be [causally] consequential it must be recognized as such, interpreted, cognitively grasped, by a consciousness' (1991: 129). This view is partly backed by her invocation of Gibson's psychology (for which, in the context of memory, see chapter 15 below), and is intended as a critique not only of Descartes' mistakes but also of those which 'permeate contemporary materialist accounts in epistemology and cognitive psychology' (1991: 130):[28] but it entails that there could

27 On 'second-order physical properties' in a similar context see Fodor 1985b/1991: 41–2. Fodor is talking of 'syntax' but patterns and dispositions fit too. Richard Carter's insightful account of three uses of the concept of disposition in Descartes includes the interrelation of parts (1983: 99–103; see also Krell 1990: 320–1, n.13). The related notions of texture and power also played important parts in subsequent developments, especially in Locke: see chapter 7 below. Dennett (1991b) defends (a form of) realism about patterns and the legitimacy of invoking them in ontology and cognitive psychology.

28 Earlier Grosholz singles out the Churchlands and Quine as representatives of those modern philosophies which bear 'the traces of Descartes' reductive method' and which support a 'hegemony of the notion of theory' which she vehemently opposes (1991: 10). But neither reduction nor theory entails, as she assumes, the ruling out of complexity of the 'multifarious' and 'prodigal' kind (1991: 81, 9–10) which she wants to defend. Reductionism is not eliminativism, and does not require the denial of the multiplicity of levels of description (1991: 127). Where Grosholz convicts Descartes of illegitimately violating his own order of reasons by bringing 'complexity in by the back door' (1991: 72), I see mechanistic openness to complexity as the way to imagine, feel, and understand it from underneath, distrusting assumptions that real complexity at a level requires that level to be irreducible to or autonomous of other physical processes. For a reply to similar criticisms see Paul Churchland 1989b.

be no patterns in nature without consciousnesses to grasp and interpret them. This is one interpretation of the Gibsonian stress on the existence of environmental regularities *only* as affordances for the active organism: but it seems an unnecessary slide away from realism about the independent existence of patterns. The fact that the animal spirits, 'almost always differing among themselves in some way' (AT xi.180, H 91; see Foti 1986a: 77–8 and Krell 1990: 65–8 on the uses of 'difference' in *L'Homme*), actualise or embody patterns of motions and thus mediate in intelligent processing, does not entail that they are themselves required to be intelligent.

3.5.2 *Animated statues, liberty, and psychological control*

If I am right, how pure or strict can Cartesian mechanism ever have been? Were the machines taken as models for natural bodies *not* then passive and static? There are different forms of this question, which I seek to clarify in this section by showing how the active/passive boundary shifts (compare Lloyd 1993a: 78–82). Cartesian automata and body-machines did, contrary to common views, have their own 'activity', in a certain (deterministic and temporary) independence from external control. They did not, however, have the acausal autonomy attributed to souls which will and act freely, and judge rationally. From the metaphysical perspective in which only the immaterial can be truly active, of course body-machines are passive. But it is appropriate still to use 'activity' of them in a Cartesian world, because of their resistance to the commands of the will, the inability of the rational soul perfectly to control and order their movements or clearly to understand their state.

Again, I take as stalking-horse a firm, representative statement of an interpretation different from my own. Otto Mayr (1986: xviii) aligns Descartes with the 'growing commitment on the Continent to the value of authority' against an opposing commitment 'in England, to its antithesis, liberty'. The mechanical philosophers relied on the metaphor of the clock for natural philosophy because of its obvious reliance on external regulation (1986: 28–121; compare Merchant 1980: 217–27). Only slowly did English theorists question the authoritarian implications of mechanism and look to the coincident development of feedback systems in technology for alternative models of balance, harmony, and *self*-regulation (without external control) which better suited a more democratic 'liberal conception of order' (Mayr 1980, 1986: 122–99). Mayr's account is much more nuanced than this: but even if his strict dichotomy between authority and liberty could be maintained, his scheme is the wrong way round. In chapters 5 and 9 I argue that English natural philosophers rejected the Cartesian philosophy of the brain not because of its authoritarian nature but because it was not authoritarian *enough*: *too much* liberty was given by Descartes to the confusing animal spirits, which had to be

stabilised, neutralised, and finally eliminated in the search for some control over memory and the past.

On the 'Continental' side, Mayr says that the physiological scheme of *L'Homme* assumes an analogously hierarchical organisation in simple machines and in organisms, in which a 'single supreme organ' has sovereign power over the system (1986: 62–7). In humans the incorporeal soul rules and regulates, but

> in animals it was the brain that, consisting entirely of memory and therefore capable only of initiating preprogrammed action, corresponded closely to the mechanical program controlling automata. Between the different levels of the hierarchy, organs were connected in linear, unidirectional, cause-and-effect, or command chains. (Mayr 1986: 66)

There are insights in this misreading. In a sense Cartesian brains do consist entirely of memory (except for the pineal gland which has to be kept clear of memory motions); and there is a central executive organ, independent in body-machines of the soul (though it is the pineal gland rather than the entire brain). But, as I argued against Flanagan above, brains (or other machines) 'consisting of' memory are *not* thereby restricted to pre-programmed action. Memory goes with *learning*. It is precisely memory which, carrying particular contingent experience in its folds, guarantees the specificity of the organism's behaviour and cognition in a context. Corporeal memory provides for the machine a flexibility or an independence from the arbitrariness of current stimuli, not in the sense of spontaneity or indifference, but in the sense that the constraints which act on the body, the ties that bind, are constraints and ties to that particular body, to that particular past.

Indeed Mayr's own more detailed analysis undermines his meta-narrative. In his official view, mechanism is a system with a 'single central cause' (whether God, king, soul, or brain), where this centre has command of all information, memory, judgement, decision, and spontaneous action (1986: 117–18). But this does not in fact apply, as his story requires it must, either to Descartes or to clocks. Mayr is puzzled that the program function imposed to guide 'program-controlled' Cartesian automata was 'not located in a single, easily-identifiable element but was distributed throughout the mechanism', in, for example, 'the disposition of the organs' (1986: 66): he is forced to complain in a note (1986: 219–20, n. 61) that Descartes does not exploit his own hints about cooperative interaction between systems which could allow for unprogrammed responses to unrehearsed situations. But in fact Descartes, like other mechanistic physiologists, is fully aware of the importance of feedback and cooperation in the bodily 'maintenance system': digestive, respiratory, and circulatory homeostasis is necessary for the reliable operation of the cognitive-informational system (Rorty 1992:

377–80).[29] Memory is not located (wholly) at the centre but is distributed across the brain surfaces and out on the independent bodily periphery. So body-machines are not narrow, specialised slaves as Mayr claims: they lack the soul's freedom, but they do not lack flexibility, a degree of self-maintenance, and adaptiveness to new environments.

In turn, the clock metaphor, meant to carry the case that mechanism and authoritarianism are intimate, is more complicated. Mayr's mechanists took clocks, like mechanical organisms, to have a single central origin governing ordered passive parts in a hierarchy. But he acknowledges that there was no consensus about what the central executive was: the original designer, the energy source, the regulator, the memory-analogue, or what (1986: 42, 117)? Georges Canguilhem, in contrast, used Descartes' *Description of the Human Body* (AT xi.225, CSM 1.315) to reject the simple authoritarian interpretation. The body is not commanded by a sovereign soul:

> Envisioning the body in terms of a clockwork mechanism, Descartes saw the various organs as controlling one another in much the same way as the gears of a clock. Descartes thus replaced a political image of command and a magical type of causality (involving words or signs) with a technological image of 'control' and a positive type of causality involving a mechanical meshing or linkage. (Canguilhem 1952/1994: 231)

Two different strains of properties are being attributed to machines and thence to mechanism (as seen also by Baker and Morris 1996: 91–4, who develop the point differently). Mayr's official line stresses only machines' dull mechanical nature: they are passive, entirely predictable, unable to give rise to complexity. But as well as dullness and passivity, machines seem unreliable, dangerous, uncertain, likely to escape the control of their designers. This is not an invention of modern science fiction: the uncanny capacity of imagined and dreamed machines to mimic and then exceed human possibility is a recurrent theme in historical fantasies of automata between technology and natural magic. Mayr knows this, detailing not only traditional tales of quasi-demonic automata (1986: 24–6; compare de Solla Price 1964: 10–12; and Bedini 1964), but also early modern references to the fragility and untrustworthiness of machines (1986: 42–53, 124). But because he attributes only to the English a 'rejection of the clock metaphor in the name of liberty' (1986: 122), Mayr fails to see that such concerns about the complexity and unreliability of mechanisms

29 Amelie Rorty in this fine paper demonstrates just how intertwined epistemology and physiological homeostasis are in Descartes' approach to 'thinking with the body'. My only query is whether it is possible to divide informational and maintenance systems even to the extent she does. She does not address the importance of the specific point that animal spirits, the bearers of distorted or accurate information, are themselves generated in and marked by non-cognitive bodily processes. See below (section 3.5.3) on the permeable body.

spring from genuine dynamic properties of the (Continental) mechanical models themselves.

If machines and mechanical models had only the properties of dullness and passivity, their subordination to external authority (God and soul) and the predictability of their behaviour would not have been questioned. There would seem no risks of confusion and danger. The fact that both proponents and critics of mechanism were in fact all too aware of the threat of ataxy and disorder which machines could bring suggests that there may have been some theoretical basis for the perceived, problematic activity. The dynamic elements in mechanical models made it seem possible that machines might *not* be wholly and obediently subject to the demands and commands of external authority. In body-machines, the psychological control which (many different) official views of soul, will, and reason required seemed menaced by the ongoing physical processes by which fibres, pores, spirits, and internal organs 'ticked' or flowed on, often inaccessible to consciousness and thus unpredictable.

Carolyn Merchant usefully separates two aspects of the 'reordering of reality' achieved by mechanism. 'Autonomous' machines like clocks symbolised *order*, while non-autonomous technology like windmills and pumps symbolised human *power*: 'Order was attained through an emphasis on the motion of indivisible parts subject to mathematical laws and the rejection of unpredictable animistic sources of change. Power was achieved through immediate active intervention in a secularized world' (Merchant 1980: 216–17). Different emphases on order or power depended partly on a position about God, whether intellectualist (God's ordered rational intellect over His will) or voluntarist (God's supremely free power of the will, unbound by logic or law: see Ayers 1993). Merchant's 'order' pole corresponds to my category of dull passive mechanical machines. But her 'power' pole needs to be further divided. Of course machines were meant to reveal the operative power of God and of the soul, the teleological need for a designer and for external governance of predictable, ordered nature and the passive body. But in practice, menacing the rhetoric of control, machines were often invoked in language or in contexts where it is their power, and their threat to incorporeal power, which seems central.

Writing in the late sixteenth century, Thomas Nashe, in *The Unfortunate Traveller*, describes fictional songbird-automata ('sweet resembled substances without sense') which produce wondrous harmonies due to fluid flows through the 'many edged unsundered writhings & crankled wanderings aside' in 'long silver pipes enrinded in the entrails of the boughs whereon they sat' (in Knoespel 1992: 113). There is nothing straightforward in the way even simple animal-machines are seen, no easy relegation to the passive. Charged metaphorical descriptions of machines which really are pre-programmed, which have no capacity for learning and memory, suggest some of the cultural weight imbued in the more complex, more dangerous cyborgs.

This suggests just how appropriate was the strange machine analogy used by Descartes to illuminate his account of the body-machine: he compares the animated statues of his fable in *L'Homme* to finely wrought devices in 'the gardens of our kings'.

> External objects which merely by their presence act on the organs of sense and by this means force them to move in several different ways, depending on how the parts of the brain are arranged, are like strangers who, entering some of the grottoes of these fountains, unwittingly cause the movements that then occur, since they cannot enter without stepping on certain tiles so arranged that, for example, if they approach a Diana bathing they will cause her to hide in the reeds; and if they pass farther to pursue her they will cause a Neptune to advance and menace them with his trident; or if they go in another direction they will make a marine monster come out and spew water in their faces, or other such things according to the whims of the engineers who made them. (AT xi.130–2, H 21–2)

The drama of the royal garden grotto (de Caus 1659; Jaynes 1970), argue critics of mechanism, has nothing to do with the simple hydraulic mechanism beneath the disguise. Wells (1985: 80) complains, for example, at the subtle suggestion that *L'Homme*'s austere explanatory apparatus of spirits, fluids, and nerves could, when scaled up, give rise to such affective complexity. But the analogy has a certain strangeness in its detail.

Most of the analogical elements are easy enough to apply to *L'Homme*'s animated statues. The nerves are the fountains' tubes for carrying water, the ventricles are the water main, muscles and tendons are the 'engines and springs', and so on (animal spirits as water, heart as water source, God as engineer). But the analogy vanishes at a key point. There is no obvious analogue for Diana and her marine defenders Neptune and the monster. They can only be equivalent to particular internal parts of the body: violent reprisal is potentially internal in origin, as the body carries destructive capacities.[30] This motivates the Cartesian ethics of knowledge of, and control over, physiology. Philosophers' alleged subduing, murder, and forgetting of the human body, their neglect of the microprocesses of human nature, was not a significant seventeenth-century phenomenon. There was continual awareness of and vigilance towards the violent, excessive body, which through early modern philosophy was always being urgently rethought.

30 Erica Harth (1992: 95–8) quotes an intriguing poem of the 1690s by Catherine Descartes, the philosopher's niece, which reinterprets Diana as an angry Nature turning on the dying Descartes. Harth sees Catherine (re)infusing Nature with life and soul, seeking to reverse the mechanists' murder of nature, who 'kills Descartes' discourse by feminising it' (1992: 98). But she does not note the additional direct parodic echo of *L'Homme* in Catherine's verse: Descartes dying in Sweden is the peeping Actaeon being turned on by his own body-monsters. (Renaissance interpretations of the Actaeon myth took Actaeon's hounds to represent his own thoughts and desires (Barkan 1980; Vickers 1982).)

As in Descartes' other flirtations with automata (de Solla Price 1964: 22–3; Rodis-Lewis 1978: 154–9; Shea 1991: 107–9; Gaukroger 1995: 1–3, 63–4), the emotive blurring of myth, magic, and mechanism in the Diana analogy reveals the need for both technology and theory to retain contact with 'the exigencies of life . . . since we cannot make ourselves a new body, we must augment our internal organs with external organs and supplement our natural ones with artificial ones' (Canguilhem 1937/1994: 225). It is just because body-machines are weak and exist in history, because hair turns white (AT i.435), that medicine is central to Cartesian philosophy. The myth of the pre-programmed machine dully reproducing its hardwired fate, eternally churning out fixed action patterns, does not apply to the animated statues of organic nature. The intermediate level of response, neither simple reflex nor incorporeally derived action, opens up Cartesian bodies to memory and history, with all the sadness, resistance, and complexity which the matter of the past brings along.

3.5.3 The permeable body

A damaging aspect of the 'the submergence of the organism by the machine' (Merchant 1980: 193), which many attribute to Descartes and his ilk, was the closing off of the human body from the world, rendering it a possession of the individual soul. Descartes is an easy villain in New Age psychosomatics and 'liberatory eco-holism' (see Brown 1985). Similarly, historians caught up by Descartes' metaphysics of free active mind gesture at critical mistakes made in the scientistic physiological construction of the passive body. Descartes sought to reduce all bodies to sameness, to fit a single micromechanical model, imagining them as automata 'endlessly repeatable, and by definition not particular, not the subjects of a specific history' (Reiss 1996: 604): they can be manipulated by the interventionist scientist just as they are by the immaterial soul, and are distant from the rational mastery exercised by this immaterial governor. Bodily events, including death, occur as if to another: 'the true self cannot be threatened by the demise of that which from the start was mere mechanism . . . The corporeal threat is, as far as possible, subdued' (Leder 1990: 148).

I hope it is now obvious that this picture is misleading. It is not only that the body's homeostatic 'maintenance system' must remain healthy to ground reliable information processing (Rorty 1992; Mackenzie 1989: 173–5). In fact, the Cartesian 'body, with its interactive openness', far from being inevitably moulded to one hardwired model, is the means by which difference is introduced into the human compound (Foti 1986a: 76; Reiss 1996). Particular experience teaches through the mixed dispositional traces enfolded in brain and body which, when actualised, cause action in a specific context.

In the ancient and Renaissance physiologies of humours and spirits, across boundaries between Aristotelian and Hippocratic/Galenic systems, the body was by nature open, the internal environment always in dynamic interrelation

with the external environment (chapter 2 above). Its state depended on the non-naturals, on regulation of temperature, and on the maintenance of fragile fluid balances in the internal media. Certain proper mean states could maintain its normal capacities to resist the immediate stimuli and to avoid surrendering its health/balance in temporary environmental upheavals. Urgent steps could be taken to close off its vents and windows, barring the orifices by which external dangers could intrude. But this seasonal body was always vulnerable to climatic effects, and permeated by the environment right through to its cognitive capacities.

Almost all of this survives in Descartes' 'corpuscularized Galenism' (Grosholz 1991: 120), transformed into principles of fluid mechanics by which inner and outer continually interact. A long passage in *L'Homme* (AT xi.167–70, H 73–5) ends: 'in sum, whatever can cause any change in the blood can also cause change in the spirits'. Here the animal spirits, as ever, are the medium of interplay between body and world. Descartes retains the central role given to blood in traditional physiology, and examines the internal and external variables affecting the spirits derived from the blood, including food and digestion (the body's nimbleness and quickness depends on the qualities of food), respiration and climate, and the states of liver, gall bladder, spleen, and heart. There are lines of causal influence straight from the non-naturals through blood and spirits to the quality and context-specific nature of cognitive functions (given that, as I have argued, animals and other 'machines' have such functions even without a soul).

The model for these processes is a kind of circulation or *antiperistasis*, obvious when the body exists in a plenum, and giving further content to the notion of holism in Descartes' physiology: everything affects everything else, in the body as in the cosmos (Carter 1983, 1985/1991). The causal factors affecting the animal spirits and thus all psychophysiological processes are numerous, and certainly do not stop at the limits of the skin. As Malebranche would state bluntly, introducing his account of the passions, the Cartesian view of the body implies that 'we are to some extent joined to the entire universe': because of the nature of animal spirits, culture and cosmos permeate the innards, for everyone is joined 'through his body to his relatives, friends, city, prince, country, clothes, house, land, horse, dog, to the entire earth, the sun, the stars, to all the heavens' (Malebranche, LO 342, v.2; compare Sutton 1998). Medical knowledge and power will be prescriptions for a regimen to cope with environmental shifts, seeking to ensure that the body can again be appropriately coupled to its physical and social surroundings.

In spelling out optimistically the explanatory scope of spirits theory, Descartes uses the language of humours (*L'Homme* AT xi.166–7, H 72–3). The mechanisation of physiology is a reduction of Galenic humoralism rather than an elimination (compare A 438). Further, there are even echoes of the

Renaissance assumption that body fluids are all in principle interconvertible. Blood, milk, fat, sperm, spirits, humours, sweat, tears, and so on, were all versions of the same set of fluids, all subject to the same principles of fluid mechanics (Laqueur 1990: 35–43, 103–8). Only in this light does Descartes' insistence on the common source of animal spirits and generative seed make sense (AT xi.128, H 18; chapter 9 below).

It is not that Descartes' mechanisation in these domains is incomplete, but that a physiology modelled on hydrodynamics explicitly theorises an active, runny, permeable body, set in a full fluid universe. Descartes' moral advice on psychological control is a set of maxims for trying to bind this open body, to stabilise the flux. The strangeness of this demand is worth pausing on. It should complicate further the difficult attempt to document multifaceted conceptual and phenomenological shifts from grotesque and open to classical and closed bodies, from spectacular to docile bodies, or from public to private bodies (Bakhtin 1965/1984: Foucault 1977: 16–31; Barker 1984). Theory which itself is alien to us imposes puzzling requirements for self-control, in which the part of nature that it is most important to master is the part we might have thought we already possessed, our own fluid spirits and the unpredictable body bits through which they pass.

3.6 Objection 4: memory, method, and metaphysics

A final objection to my interpretation of Descartes' remarks on memory lurks in the puzzling connections between the physiology of memory and Cartesian metaphysics. The destruction of opinion is, notoriously, the prerequisite of Cartesian method. A new foundation for thought is possible once the house of belief has been rased. The natural light of reason will suffice as origin for a new structure. It seems as if the method, then, requires the total forgetting of everything learned and taken on trust, the effacing of book 'knowledge' from the mind. We must, as Andrew Benjamin (1993: 42) puts it, refuse the gift of the given, closing off the possibility of repetition in order, like Descartes, to take up our subject-matter as if for the first time (compare Schouls 1989: 144–5 on Descartes' distrust of the past).

This demand that all opinion be erased is provocatively discussed in the context of Descartes' theories of memory by Benjamin and by Tim Reiss. The difficulty is that the idea of *deciding* to forget is not obviously coherent. Reiss (1996: 595) quotes Gassendi's complaint that 'memory . . . cannot be erased at will'. Even if it could, a regress of forgetting looms: the memory of the act of erasure would itself conceal the old opinion, and would thus itself have to be forgotten, *ad infinitum* (Benjamin 1993: 50–1).

Benjamin leaves the dilemma in place in characterising the 'inherent instability' of forgetting: either the regress ensues, or the old opinion remains

traced in the memory alongside the new, when the inadvertent intrusion of unwanted traces into the new rational thought will always remain a threat. But Reiss thinks that Descartes answered Gassendi's challenge: the emptying of memory is not, in fact, required. After subjecting beliefs to rational examination, one will know by the natural light which candidate beliefs to assent to: but even after such rational rejection of old ideas, 'one does not fail, for all that, to retain the same notions in one's memory' (to Clerselier, 12 January 1646, AT ix.204, in Reiss 1996: 596). The positive procedure for evaluating beliefs is to balance old ideas against new, by multiplying contrary traces: critics like Gassendi missed this precise project of avoiding

> those many divers traces always already imprinted on, bored through, folded into the brain. They did not grasp that it was never a question of making these vestiges disappear. Descartes agreed that such a project would have been time wasted, because impossible. The body's marks were indelible. What he sought, rather, was to render this physical history inoperative. (Reiss 1996: 601)

Reiss' claim that Descartes did not, ludicrously, believe in the self's perfect control over memory is exactly right. It is rare for us really to have a hand in our forgetting. But if Reiss is correct about Descartes' reasons for denying this freedom to forget, then my account of distributed memory in Descartes cannot be right. Reiss (1996: 597–9) analyses Descartes' descriptions of the physiology of memory, from the impressions in wax of the *Regulae*, through the stories in *L'Homme* of nervous spirits ('what we would now describe as neuro-electrical impulses') opening holes in the tissue of the brain, to the discussions of memory folds in the correspondence. In all these models of memory traces, 'one thing was sure. Whether imprint, hole, or fold, the impressed mark was permanent. It was there, once and for all' (Reiss 1996: 599–600).

By contrast, according to my interpretation, no marks are permanent (save perhaps the rare traces deeply graven in excessive wonder): dispositional traces continually act together, and are not even easily isolable. Permanence, possible only if memory representations remain explicit throughout the storage period, is impossible if many traces are stored in the same physical system: if remembering requires the reconstruction and re-separation of memory motions in the animal spirits, then interference effects make it unlikely that traces of old opinions will be formed perfectly identical to those at the time of initial learning.

Descartes does not state that traces are permanent and unchanging. When he writes to Clerselier that after rational examination of an old and acceptance of a new belief, one retains the same notions in one's memory, this does not, I suggest, mean that items in memory never change: it means, rather, that they do not disappear *at will*, that the act of rational examination may not itself change them.

Reiss' central contention about our lack of control over memory can be maintained without committing Descartes, implausibly, to the permanence of items in memory. The threat of continuous mixture in memory was more worrying than any model which guaranteed permanence, and this is one reason why Descartes is so concerned, when thinking about method, to finds ways of bypassing memory entirely.[31] How can truth be preserved in thinking when memories are superpositionally stored in the folds of the brain, blended or confused one with another? Reiss also suggests that Descartes was concerned that memory is unreliable, weak, and unstable (Reiss 1996: 600), always bringing obscure and confused ideas to mind, mixing impressions derived from different sources, adding prejudice to perception. Although either permanence or unreliability would explain why memory and forgetting are beyond conscious control and free decision, it seems unlikely that memory traces are both unstable and permanent. I suggest that the tension here can be resolved by accepting that the unreliability, not the permanence, was Descartes' worry. It is not that traces are too static, mechanical, and dully permanent, but that they are too fickle, fleeting, and fragile.[32]

Both Benjamin and Reiss note the irredeemably bodily nature of the constraints which memory imposes on Cartesian method. The soul never has absolute power over corporeal things. But bodies are in motion incessantly, not forever fixed.

3.7 Control and self-mastery

The four objections do not succeed in refuting the idea that Descartes had an inchoate distributed model of memory. But there is still good reason to classify his theory of memory as not just historically but also conceptually distinct from the modern connectionist picture. Descartes did acknowledge that non-conscious memory processing occurs out in the folds of the brain, or elsewhere in the body: this looks, as I have suggested, not unlike the

31 This is most explicit in the *Regulae*, where the intellect relies on memory to connect the steps in a reasoning process. But the instability of memory means that the task of reason is to run over all the steps of an argument until 'memory is left with practically no role to play' because intuition can encompass the whole (Rule 11, AT x.409, CSM 1.38). This relieves our memory and enlarges intellectual capacity (Rule 7, AT x.387–8, CSM 1.35).

32 One of Tim Reiss' suggestions in correspondence is that Descartes' problem about memory and method is not so much about how to neutralise fixed memory traces, but about how to render them dynamic, and thus open to rational re-evaluation. As Reiss puts it, 'that memories need reconstructing does not contradict the fact that they can't be erased': but he wonders whether, on my interpretation 'reconstruction can be the right term, for, if so, we're surely no longer talking with any security about *memory*? Are we not, then, in the realm of either imagination or judgment?' (personal communication). This exactly is Descartes' worry, for, on my interpretation, traces already are dynamic, but tending more in the direction of imagination, beyond the pale of reason and will, than of judgement.

connectionist urge to collapse storage and processing so that computation is incessant rather than occasional. But Descartes retains a gap between storage and processing to the extent that he still postulates a central executive mechanism located at the pineal gland, which must be kept clean for vigorous ongoing processing of input and computation. Not only beings with rational souls, but also Cartesian automata, have this singular control engine through which animal spirits must be filtered and across which corporeal ideas play.

We could ascribe the role of this control centre entirely to Descartes' metaphysical dualism, and then stress the highly restricted scope of the incorporeal in the physical world. Yet however minimal the dualism, Descartes still imposes a single centre which receives input from body and world and, in its swivelling and jiggling, orients the organism's responses (Dennett 1991a). Descartes' worries about control are directly connected with memory. In conversation with Burman, Descartes said that 'the difference between sense-perception and imagination is thus really just this, that in sense-perception the images are imprinted by external objects which are actually present, whilst in imagination the images are imprinted by the mind without any external objects and with the windows shut, as it were' (AT v.162–3, CSM-K 145). Memory, significantly, just does not fit this scheme, for it bridges the distinction between untrustworthy constructive imagination and passive input-driven perception. In memory, objects or events are absent, for they are in the past. But memory, imagination, and sensation are all available to non-human animals and to humans when their minds are elsewhere. So the 'painting' of representations in the spirits can go on without being done by the mind, and without being directed by the objective external world. Who, or what, is in charge? There is only the folding, sieving, and commotion of the physiological fluids and fibres.

It is for this reason that Descartes and, more explicitly, Malebranche (chapter 9 below) construct strategies of evasion, ways of reasserting control over the wayward body. Longing for inner discipline springs, in part, from awareness of the dangers of distributed memory: and the fragile authority of a central executive, whether the incorporeal intellectual memory or the rolling pineal gland, is the only security.

Some seventeenth-century memory theorists, like Digby and Hooke (chapter 5), favoured local models of memory just because they required a simple picture of strong central control of cognition, with passive items needing active manipulation. But distributed models, which take traces to be in, or to be, the interrelations of physical parts, do not sit so easily with this ideal of global control, of an executive with access to and control over what is going on everywhere in the system. On a distributed model, memories actively alter all the time.

Where then does Descartes slot in? The history looks paradoxical. I have argued that Descartes had at least a partial understanding of distributed models: but, notoriously, Descartes posits a central executive, both for dualist purposes and as a control mechanism. Something must give: either Descartes really does not have a distributed model, or distributed models can in fact coexist with a central executive, or Descartes does not have a central executive. This is the key historical crux in Descartes, the point at which coherence gives out. There is much to be said for the second option, that distributed memories can be made compatible with central control, for the local versus distributed distinction is technically independent of issues of cognitive control. But in Cartesian terms, with regard to the work of psychological explanation, Descartes does not in effect have much of an executive, the pineal's control and anchoring functions being swamped by the physiological dynamics: his theory is under terrible internal pressure.

The wishful executive does not have access to the whole system, to all the information passing through the pores, for the continually shifting fluids which ground memory capacities cannot always all be tracing their figures on the surface of the gland. They whoosh round in patterns through brain and body, causing all sorts of motions of which the executive has no inkling. The pineal gland is buffeted on all sides, beset around by the winds and flames of the animal spirits. Hence Descartes' increasing attention to psychological conflict, and hence his recurring desire to distance himself from his own theories of physical-cognitive processes like remembering by mentioning piously the importance of intellectual memory, or by seeking to bypass and eliminate memory from the life of the rational agent. Executive control is either illusory, or, at best, fragile.

Yet there is a further, positive picture of self-mastery, which holds both control and memory activity together, just visible in Descartes' later work and in Malebranche. The plasticity of corporeal memory is indeed dangerous: but it is, also, the soul's only hope. Moral neurology was not the old dualist diatribe against the body, but the knowing use of habit and association to encourage the sources of passion in brain and body to shift into morally sanctioned paths. The *Passions*' recommendation to understand physiology in order to live better is not just scientistic raving, requiring direct introspection of our brain states. It involves provisional maxims, applicable differently in each individual, for applying intelligence to the reflexes, and (fallibly, interminably) re-colonising the body.[33]

Though I cannot expand on this suggestion here, it is worth pointing out that the much maligned Cartesian gap between self and body starts to look better

33 Thanks to Doris McIlwain for suggesting this way of formulating my interpretation. It is developed a little further in Sutton 1998: compare Morgan 1994: 154.

motivated from this perspective. John Cook (1969: 118–21) complained that Descartes 'failed to realize that he was introducing an extraordinary sense of the word "body"', in which a 'distinction between *himself* and his body' makes sense (where it allegedly does not in ordinary language). But the programme of extended autopersuasion which Descartes advised for Princess Elizabeth allowed the active mind to mould associative responses, gradually becoming an architect of one's own passions and (correspondingly) of the landscape of pores and fibres which the spirits sculpt. Against the immediate dictates of the preservation of the body, the task is to work towards the permeation of body and brain by intellect and will, a permeation directly parallel with, and often contrary to, that already enacted by the physical world on the body. As thinking beings embedded in living body-machines, we must often correct for the hasty norms appropriate for those machines as biological beings alone. Only thus might the compound creature which thinks and eats, reasons and dreams in such a marked, particular body ever become more truly what, as a unified whole, it is.

Appendix 1: nerves, spirits, and traces in Descartes

I have suggested a new interpretation of Descartes' views on corporeal memory, and shown how it was embedded in his general physiology and natural philosophy. In this appendix, to set memory even more firmly in Cartesian nature, I provide further background to the theoretical framework of *L'Homme* and Cartesian physiology for readers unfamiliar with it (Kemp Smith 1952: ch. 5; Riese 1958; Hall 1969, 1970; Krell 1990: 62–73; Grosholz 1991: ch. 6; Shea 1991: 182–9; Gaukroger 1995: 269–90). The structure of *L'Homme* repays close attention: it is a complex weave of postponements, repetitions, promises, and reconstructions which often comes at the same topics from a number of different angles.[34]

The machine Descartes describes 'can be moved in all the ways that our body can, solely by the force of the animal spirits which flow from the brain into the nerves' (*L'Homme*, AT xi.137, H 29). He departs from neurological tradition only in placing sensory and motor functions in the same nerve, in opposition to both Galen and Vesalius (AT xi.133–5, H 22–6 and notes). The nerves are hollow tubules or pipes enclosed in a double membrane continuous with the brain's *pia mater* and *dura mater* (AT xi.133, H 23; on hollow nerves see Clarke 1968; Clarke and Jacyna 1987: 160f.). They contain threads or tiny fibres in a central marrow running between the brain and the periphery; the fibres are sur-

34 Hall's judgement that *L'Homme* is 'topically unbalanced' presupposes that it was intended as a general physiological primer, with completeness as a goal: he is closer in describing it as 'an essay in physiological psychology with supportive – but subordinate – sections on other physiological topics' (H xxxvii).

rounded and protected by animal spirits (AT xi.134, H 24; AT xi.141–3, H 33–6, CSM I.101–2).[35]

Theories of sensation and neuromuscular action follow. In sensation, stimuli displace the peripheral and thus, simultaneously, the central ends of the nerve fibres, instantaneously and mechanically transmitting to the brain some kind of structural isomorph of the impression made on the sense organ (AT xi.141–5, H 33–40 and notes, CSM I.101–3; *Passions* I.12–14, AT xi.337–9, CSM I.332–4; Brazier 1984: 18–29; Clarke 1989: 27f.).[36] The mechanisms by which different sense organs transmit patterns to the brain are discussed at length. Blows on the tympanic membrane from the outside air, for instance, pass to the brain by the nerves, and the soul conceives the idea of different sounds depending on the frequency, and on various relations between the blows. Quantitative differences between patterns of blows result in qualitatively different sensations of hearing (AT xi.149–51, H 45–8; on music, hearing, and harmony see Shea 1991: 69–92; Gaukroger 1995: 74–80, 286–7; Kassler 1995: 43–8). The isomorph, which is a set of physical states of the brain resulting from the transmission of motions from the organs, somehow makes the soul have different sensations (AT xi.143–4, H 36–40, CSM I.102–3). Resultant changes in the patterns of motions of spirit particles in the brain in turn cause differential outflows of the spirits into the nerves (AT. xi.159–60, 171, H 59–60, 77–8), causing muscular motion by inflation (Nayler 1993: 38–89).

Descartes eliminates the traditional 'vital' spirits:[37] animal spirits pre-exist as the finest, most subtle, and fastest moving vigorous parts of the blood. These are 'extremely small bodies which move very quickly, like the jets of flame that come from a torch. They never stop in any place' (*Passions* I.10, AT xi.335, CSM

35 At *Dioptrics* IV (AT vi.109–12) Descartes reiterates the three components of the nerves: enclosing membranes, internal threads, and animal spirits. Although 'the anatomists and doctors have adequately demonstrated that these three things are in the nerves . . . no one among them has adequately distinguished their uses'.

36 The internal nerve marrow is involved in this sensory process only, and has no motor function. The animal spirits transmit only motor impulses all the way between brain and periphery. But they are still necessary for sensation. It is not only that their function in inflating the nerve tubes ensures that the small fibres which run from periphery to brain 'do not crowd or impede each other in any way' (*Dioptrics* IV, AT vi.110); they also, crucially, must receive in their patterned motions through the brain, a sensory representation of the cause of the sensory input. This is why disruptions in the state of the spirits, a proneness to turbulence or stagnation, can threaten confidence in the subject's comprehension of, as well as interaction with, the world. Starobinski (1966: 176–7) takes a different view, that Descartes privileges muscular motion over sensation.

37 Descartes accepts the existence of particles drawn out of food which others call natural spirits and of particles heated by the heart and separated from other parts of the blood, 'which the medical men call "vital spirits"'. But 'there is virtually no difference' between these two, and they are unlike the 'pure' animal spirits in not being separated from the blood (to Vorstius, 19 June 1643, AT iii.688–9, CSM-K 225–6). As Hall says (H 21, n. 43), Descartes gives these traditional spirits no role of their own, and discards them from actual physiological explanation.

1.331–2). After their separation from the blood and passage through the pineal gland, they enter the brain and its unstable economy of spirits. They flow through the pores of the brain, to which Descartes assigns a central explanatory role. These are 'no different from the spaces that occur between the threads of some tissue; because, in effect, the whole brain is nothing but a tissue constituted in a particular way' (*L'Homme*, AT xi.170, H 77: on the term 'tissue' see Hall's note 123 (H 77) and Rather 1982: 138–40).

The brain is 'a rather dense and compact net or mesh' linked by 'conduits' which receive the animal spirits as they leave the pineal gland via the ventricles (AT xi.171, H 77–8). Threads arise from this net and extend to the rest of the body as the bundles of filaments enclosed within nerve membranes. These are channels, as Hall puts it, which conduct animal spirits into the nerves with which they are continuous (AT xi.171, H 78, n. 124).

The whole brain mesh, composed of filaments with pores between them, is affected by the particular motions of the animal spirits in three ways. The pores can be 'diversely enlarged or constricted by the force of the spirits that enter them'. Secondly, the filaments themselves 'can be flexed rather easily' by the same force as the spirits. Thirdly, and of importance for Descartes' account of memory, the filaments 'can retain, as if made of lead or wax, the flexure last received until something exerts a contrary pressure upon them' (AT xi.171, H 79).[38]

In thus describing the pores of the brain Descartes, at this point in *L'Homme*, is outlining 'in proper order' the second of three factors on which alone life functions depend (AT xi.166, H 72, CSM 1.104). Having earlier postponed a full account of the spirits and the functions they support (AT xi.133, H 23), and moved on to deal with muscular motion, respiration, and the various external and internal senses, Descartes finally agrees to 'commence to explain to you how the animal spirits pursue their course in the cavities and pores of its [the machine's] brain, and what functions depend on them' (AT xi.165, H 71, CSM 1.103). Just as the harmonies of a church organ depend not just on the structure of its parts, but on the supply and distribution of air through its pipes, so the harmonious functioning of the body-machine depends not on gross structural anatomy but on (a) the animal spirits; (b) 'the pores of the brain through which they pass'; and (c) 'the way in which these spirits are distributed to these pores' (AT xi.165–6, H 71–2, CSM 1.103–4). This third factor in fact marks the break-

38 Landormy (1902: 271, 265) took this to imply that the facility of brain fibres for the retention of images is just a consequence of Descartes' general principle of inertia, whereby it is a universal property of extended matter to retain a modification once applied to it. But more is required for the preservation of memory motions in the brain over longer periods of time: the problem will be to explain the possibility of reconstructing a 'flexure' after many different contrary pressures have been received, and this is what Descartes tries to do later. The idea that there is some physical constancy in the brain tissue is simply a first step.

down of the organ analogy, since it explains why organic automata are not restricted by their hardwiring to reflex responses to immediate stimuli alone.

The distribution of fluids through the brain is unceasing, since the spirits are in incessant motion (AT xi.172, H 79). They continually flow out from the pineal gland through the brain cavities, and some enter particular brain tubules.[39] Entering interstices in the fibrous mesh of brain substance, and tracing patterns by their motions through these pores, the spirits 'tend where the arrangement of the brain at the time impels them, not necessarily to regions that face them rectilineally' (AT xi.173, H 81). The specificity of patterns at a given time is important here, for this is how current context, in the form of incoming patterns of motions from the environment and from elsewhere in the body, will meet and interact with current context in the form of the existing dispositional states of the brain. Only physical variables are involved, and yet experience and individual learning history can encounter present environmental factors. The operative principles are to do with differences in motions: perceptible differences in objects are in some as yet unexplained way perceived by means of various differences in the spatiotemporal paths traced by the animal spirits in the brain.

Descartes talks of 'traces' and 'figures' as these explicit, specific patterns in the spirits at a given time.[40] The spirits flowing around the neural system keep the filaments of nerves and brain 'so tense' that patterns of motions are easily transmitted (AT xi.175, H 84). This is how 'ideas of objects that impinge on the senses' are formed: just as a figure 'corresponding to that of [the] object' is traced at the back of the eye, so 'the different ways in which tubes 2, 4, and 6 are opened trace on the internal surface of the brain a figure corresponding to that of [the object]' (AT xi.175–6, H 84–5).[41] Likewise, the pattern of motions of spirits as they leave the pineal gland traces the same figure on the gland's surface. 'Figure' here, says Descartes, refers to everything which can cause any sensations, not just to 'things that somehow represent the position of the edges and surfaces of objects' (AT xi.176, H 85). The role of 'figures' as

39 Those without sufficient strength to do so, or which are blocked by closed tubules, can pass into the nostrils and cause sneezing or be turned back on to the internal surface of the brain, causing dizziness or vertigo (H 79–80).

40 In fact, Descartes' use of these terms is more ambiguous. Such terms sometimes refer to transient patterns in or of the spirits, and sometimes to enduring changes in the brain pores. The ambiguity is not necessarily problematic, but it does invite further conceptual clarification: in chapter 6 below, I distinguish between (transient) explicit representations and (enduring) implicit representations. The distinction applies as clearly in Descartes' account of memory as in modern connectionist models.

41 Hall's note to this passage (H 85, n. 132) acknowledges that, while 'later physiologists were to repudiate animal spirits and substitute neuronal pathways within the brain for the currents envisioned by Descartes', the idea of 'a correspondence between retinal and cerebral patterns . . . was ultimately to become a stable part of the theory of visual perception'.

representations, in other words, is *not* to resemble all qualities of the object represented, but simply to vary with the perceptible qualities. Representation can operate without simple resemblance, as he stresses elsewhere (chapter 15 below).

This account provides sufficient physiological background for us to interpret the distributed model of memory which I described in section 3.2 above. Physical 'patterns' of spirit motions are the key, flowing across fibres which can retain certain flexures: in memory, certain of the superposed patterns can be formed again 'long afterward', because of the dispositional traces, in the form of physical alterations to the connectivity of pores in the brain folds.

Appendix 2: Malebranche on memory

The spread of the distributed model of memory based on patterns of motions of the animal spirits through the pores of the brain was initially due to Descartes' *Passions of the Soul* and, when published, his *L'Homme*. In chapter 5 I look at some English responses to and criticisms of these expositions. But I want to buttress the plausibility of my interpretation of Descartes through a brief consideration of the account of memory in the work of his follower Nicolas Malebranche. Malebranche's model is, I think, more explicit in its commitment to the twin requirements for distribution, the *extendedness* of a memory trace, and the *superpositional* storage of traces in the same physical system.[42]

Even more so than in Descartes' case, historical attention to Malebranche is restricted to metaphysics. There is only one reference in the *Bibliographia Malebranchiana* to memory (Easton, Lennon, and Sebba 1992: 80), namely to an Italian article of 1893 which reportedly criticises his mechanical psychophysiology of memory as incompatible with his occasionalist metaphysics, whereby there should be no continuity in the physical world. So I make no apologies for taking Malebranche's psychophysiology both seriously and out of context. Much English-language Malebranche scolarship has focused on occasionalism, on his representational theory of ideas by which we see all things in God, and on his disputes about ideas with Arnauld (Radner 1978; Nadler 1992). Yet neurophilosophy takes up a huge amount of his writing, and

42 In one of the few brief references to Malebranche's physiology by Anglophone scholars, McCracken (1983: 35), unwilling to be detained by the details of Cartesian spirits theory, wrongly takes Malebranche's views of the retentive powers of imagination to depend 'wholly on the capacity of the brain to serve as a storehouse'. Only awareness of the nature of distributed models of memory reveals just how important it was to Descartes and Malebranche that the brain was not in any obvious sense a storehouse. More general examination of 'seventeenth-century French connectionism' (Diamond 1969) will require greater attention to the psychophysiological writings of La Forge and other early Cartesians than Anglophone scholarship has yet given.

a sizeable proportion of the 5,598 uses of the word *esprit* in his major work are in the plural, referring to the *esprits animaux* (Robinet 1984).[43]

In fact, reactions to 'the Cartesian philosophy of the brain' owed as much to Malebranche's as to Descartes' version. John Wright has demonstrated, in particular, the extensive use made of Malebranche's psychophysiology by Hume (Wright 1983: 70–4, 212–15, 224–6). Six editions of Malebranche's *De la recherche de la vérité* appeared between 1674–5 and 1712, and two English translations of the fourth edition (1678) appeared during the 1690s.[44] John Yolton (1984b: 186) notes later judgements by both La Mettrie and Joseph Priestley that Malebranche was 'said to have been the first who brought into vogue the doctrine of *animal spirits*'. Malebranche received lavish praise from later psychologist-historians of theories of memory. Burnham in 1888, after lamenting Descartes' 'crude physiology' and 'dogmatism', dubbed Malebranche, for his account of brain traces and memory, 'a true pioneer in the field of physiological psychology' (1888: 66–8). Gomulicki (1953: 4), while regretting that Malebranche agreed with his contemporaries 'in still accepting such concepts as that of "animal spirits"', praises his terminology for the physiological basis of memory on the grounds that it 'might almost be modern'. As well as confirming and illustrating the Cartesian model, a brief discussion here serves as background for my use of Malebranche in chapter 9 on the relation between distributed animal spirits and views of the self.

Malebranche explicitly links his views on animal spirits and 'the structure of the brain' to the 'hypotheses or suppositions' of Descartes' *L'Homme*, suggesting that attention to it will satisfy the reader 'on all these questions because of the method he provides for their resolution' (LO 93, II.1.2.iii). Malebranche's chance reading of *L'Homme* in 1664, he related, made his heart palpitate with excitement (Heilbron 1982: 31; Gaukroger 1990: 41). But in his own exposition of the mechanical physiology of nerves and spirits, Malebranche displays a caution which is surprising within the allegedly wild rationalistic speculations allowed by Cartesian method: he reflects that proofs of his account would require 'a general survey of physics and then a very precise account of the human body' (LO 355, V.3). He is sanguine in allowing that these sciences are as yet too imprecise.

In book I, on the senses, Malebranche's 'explanation of the sense organs' is a description of their anatomy and of the physiology of sense perception. The nerve filaments running between sense organs and 'the middle of the brain' are 'hollow like little canals and are completely filled with animal spirits' which make them taut (LO 49, I.10.ii). He remains neutral over whether sensation is

43 '. . . chez lui comme chez Descartes, les conseils de méthode sont inseparables d'une conception physiologique de la mémoire, de l'imagination, et de l'association des idées' (Alquié 1974: 30).

44 By Thomas Taylor in 1694 (2nd edn, 1700), and by Richard Sault in 1694–5.

due to vibrations of the animal spirits or to 'the continuous displacement of the filaments right up to the brain' (LO 50, I.10.ii). Book II, on the imagination, is founded on the possibility of explaining 'all the different characters encountered among the minds of men' by reference to differences in the same physical variables on which Descartes relied (LO 89, II.I.I.iii). The brain fibres can be delicate or coarse, moist or dry, malleable or rigid, while the animal spirits can be abundant or scarce, rapidly or slowly agitated, dense or light, and vary in 'pressure', causing differences in 'the depth and clarity of the traces in the imagination'.[45] The variables of animal spirits and brain fibres are, almost, explanatorily exhaustive: 'it is the variety found in these two things that constitutes nearly all the great diversity observed among minds' (LO 89, II.I.I.iii).

On top of this basic explanatory apparatus,[46] Malebranche gives a mechanism for the lasting effects of experience on the brain. This, indeed, is central to the whole project of *The Search after Truth*, which is to expose and explain 'all the errors of men and their causes', to try then 'to relieve the mind of the errors to which it is subject', and to give a '*general method* for conducting the search after truth' (LO 17–18, I.4.ii–iii). The first source of error, and the subject of book I, is sense perception. Ordinary perception is when the impression made by objects on the external organs is communicated to the brain: but when 'the flow of animal spirits makes a similar impression in the brain', the soul can perceive even absent objects as represented by the imagination in the brain (LO 17, I.4.ii). Vestiges of past impressions, then, can survive. The difference between sensing and imagining is only one of degree, since the agitation of fibres in the brain is sufficient for the soul to perceive something, and such agitation can originate in internal disturbances in the flow of animal spirits as well as in impressions made by objects (LO 87–8, II.I.I.i; Wright 1983: 213;

45 The animal spirits are corporeal, we are told at LO 91 (II.I.2.i), 'merely the most refined and agitated parts of the blood', which are separated from the blood in the brain. This chapter gives the Cartesian account of the origin and formation of spirits, and II.I.3 allows for environmental effects on their nature by way of differences in the air breathed. Malebranche gives details (II.I.2.i) of the different kinds of fermentation or refinement of the blood which can affect the spirits, making them, for example, too agitated or too inactive: in addition, 'depending on the stability of the blood particles, the animal spirits will be more or less stable themselves'. In the 'examples and indisputable experiments' which Malebranche here promises to give us, this potential instability of the animal spirits is often the source of moral and psychological danger and imbalance.

46 Malebranche also accepts Descartes' hydraulic account of muscular motion, stressing that the effusion of animal spirits into muscles can be swift enough. 'The whole question', he says, 'regarding voluntary movement can be reduced to knowing how the small quantity of animal spirits contained in an arm can suddenly swell its muscles according to the orders of the will with sufficient force to lift a load of a hundred pounds and more'. Convulsive movements in turn are due to irregularities in nerves or spirits, when, for instance, the nerves are 'plugged by some humor' (LO 502–3 and 508, VI.II.8). For a discussion of Malebranche's views on action and muscular motion in the context of his occasionalist metaphysics see Yolton 1984b: 128–31 (also pp. 160–2 on Malebranche's accounts of animal spirits and brain traces).

Yolton 1984b: 160). Malebranche uses the word 'trace' for what survives in the imagination after the spirits have traced patterns in perception (LO 88–9, II.1.1.ii–iii).[47]

After a long explanation of the possible internal and external causes of changes in the animal spirits 'and as a result in the imagination' (LO 99, II.1.4.iii), Malebranche goes on in the dense chapter 5 of part I of book II to examine 'the connection between our thoughts and the brain traces, and the reciprocal connection of these traces'. This involves the same dispositional account of memory traces and of associative mechanisms for reconstructive remembering which was central to Descartes' model. The mutual connection of traces, says Malebranche, 'consists in the fact that the brain traces are so well tied to one another that none can be aroused without all those which were imprinted at the same time being aroused' (LO 105, II.1.5.ii). This is what allows for the use of names and other marks as memorial aids. Some connections between traces, like the association between 'the trace of a great elevation one sees below oneself, and from which one is in danger of falling' and the trace 'that represents death to us', are natural and 'necessary to preservation of life': such a tie between traces cannot be broken, for 'it consists in a disposition of the brain fibres that we have from birth' (LO 106, II.1.5.ii). Other connections, those which Malebranche will address at length as the source of our errors, are acquired and fortuitous, depending, for instance, on

> the *identity* of times at which they were imprinted on the brain. For it is enough that many traces were produced at the same time for them all to rise again together. This is because the animal spirits, finding the path of all the traces made at the same time half open, continue on them since it is easier for them to travel those paths than through other parts of the brain. This is the cause of memory and of the bodily habits we share with the beasts. (LO 106, II.1.5.ii)[48]

47 The patterns originally traced are usually called 'images' of objects by Malebranche, but in his first introduction of the survival of traces of absent objects, in 1.4.ii, he adds the reservation that what are formed are only 'images of them [objects], *as it were*, in the brain' (LO 17, my emphasis). There is no need to assume that Malebranche is requiring simple resemblance: the relation between trace and object is psychophysiological, at the level of isomorphism of pattern of animal spirit motions from the time of experience to the time of remembering or imagining. This too is consistent with Descartes, for whose attacks on representation as resemblance see chapter 15 below. Malebranche confirms this interpretation later when he finds it 'absolutely necessary' to remind us that 'these images are nothing other than the traces the animal spirits make in the brain' which vary in depth and clarity according only to the ease or facility with which the spirits can re-enter their paths or patterns of motions (LO 134, II.II.2). Nadler (1992: 46) stresses that 'there is not a single Cartesian in the seventeenth century who opts for the resemblance view of representation'.

48 On the distinction between natural and acquired connections see also Wright 1983: 227f.; Yolton 1984b: 162; and chapter 9 below on natural associations. The importance of Malebranche's references to what is necessary for the preservation of life is well brought out in Gaukroger 1990: 26–34.

The distributed nature of Malebranche's model appears here in that the traces do not remain, or are not conserved, explicitly from experience to retrieval. They have to 'rise again': and they can change over time as a result of continuing processing, for 'not always being necessary for the preservation of life, they need not always be the same' (LO 106, II.1.5.ii). It is in these acquired connections between traces, traces which exist only as dispositions for the animal spirits to find more easily old patterns of motion, that plasticity of mind enters Malebranche's neurophilosophy.

'The nature of memory' falls easily out of this model. It is a dispositional account in terms of the retention of a facility for the reconstruction of certain patterns of motions of spirits in fibres:

> just as the branches of a tree that have remained bent in a certain way for some time preserve a certain facility for being bent anew in the same way, so too our brain fibers, having once received certain impressions through the flow of the animal spirits and by the action of objects, retain some facility for receiving these same dispositions for some time. Now, memory consists only in this facility, since one thinks of the same things when the brain receives the same impressions. (LO 106, II.1.5.iii)[49]

The preservation of traces, then, is not a straightforward conservation of unchanging items passively stored: it consists rather in physical dispositions for the reconstruction of particular episodes of processing. The usual ambiguity within distributed models over the individuation of traces is present in this passage, most clearly in the somewhat paradoxical analogy with branches which 'preserve a certain facility for being bent anew in the same way': a distributed memory trace too is constructed anew, but in the same way as an earlier trace. Malebranche is able simply to dismiss 'the prejudice that our brain is much too small to be able to preserve such large numbers of traces and impressions' (LO 107, II.1.5.iii), for the traces do not all have to be simultaneously active: they are 'stored' only superpositionally, in 'the changes occurring in the fibers of the principal part of the brain' (LO 106, II.1.5.iii).[50]

The ease with which spirits cause associated traces to rise again together is a principal cause of the 'disorder of men's imagination' (LO 130, II.11.1), and, it becomes clear, even more so of women's. We often fail to 'judge things soundly' because 'the animal spirits ordinarily flow in the traces of the ideas most familiar to us' (LO 134, II.11.2). Because traces are distributed and super-

49 Malebranche goes on to explain 'why we do not recall all the things we have perceived equally well', and to suggest that each reader go on to work out 'the cause of all [the] surprising effects of the memory'.

50 Malebranche's confidence that finding room in the brain is no serious problem for his theory strongly confirms that it is a distributed model, and also supports my suggestion above that Descartes' view of memory did not lead him to despair of finding storage space in the brain.

posed, interference between them is inevitable: brain traces can become 'confused with each other, because there are so many of them, and reason has not arranged them in order . . . When the mind wants to open certain traces but encounters other more familiar ones crossing them, it is misled' (LO 141, II.11.4). The 'crossing' of traces in the same system is superposition: the capacity of such a system in the brain not being infinite, it is, says Malebranche, 'nearly impossible for so many traces, formed without order, to avoid becoming mixed up and bringing confusion into the ideas'. This is why remembering many things is often incompatible with ordered reason and judgement (LO 141, II.11.4).[51] The preservation of original order, then, is not a natural property of distributed memory. It is an achievement to avoid confusion, for confusion is the primitive mechanism by which remembering operates. Since memories are not stored explicitly at separate memory addresses independent of each other, generalisation, blending, and mixing are all but automatic. Traces are, as Malebranche says, 'formed without order'.

Misassociation is an ever-present danger because of this dangerous plasticity in memory. Malebranche gives a variety of 'very obvious and intuitive examples' of what can happen when 'the animal spirits, finding some resistance in the parts of the brain whence they should pass, and being easily detoured crowd into the deep traces of the ideas that are more familiar to us' (LO 135, II.11.2). We see a face or a man on horseback in the moon, we see chariots, men, or lions in the clouds, we mistake all diseases for the familiar scurvy, we dream of objects we have seen during the day, we see heads on walls which have 'many irregular colored patches', or we take everything to be connected with the subject of our continual study. Spirits of wine (which Malebranche has earlier told us are 'almost fully formed animal spirits, but libertine spirits, which do not voluntarily submit to the commands of the will' (LO 92, II.1.2.iii)) can enter easily into familiar traces and thus, without the involvement of the will, 'cause the most important secrets to be revealed' (LO 135, II.11.2).[52]

Confusions between traces, and the misassociations involved in disorders of the imagination, require much vigilance and continual efforts to assert control over the potentially libertine animal spirits. I discuss these topics in relation to the psychophysiology of self in chapter 9. Here I note remarks from later in Malebranche's work which echo Descartes on the psychophysiology of wonder, and again reveal Cartesian hesitancy about accepting the implications of the distributed model.

51 Compare the discussion of the utility of forgetting and of Montaigne's alleged excellence in forgetfulness at LO 187–8, II.111.5.

52 Wright (1983: 70) also discusses the moon example. It is, in a sense, just an extension of Descartes' example, in the original exposition in *L'Homme* of total recall on partial input, of adding all the elements of a face when we see some of them. Veridical remembering and imagining differ psychologically only in degree.

Because error is so easy on a distributed model, the kind of safe, clear, and distinct cognitions which the pure seventeenth-century moral physiologists desire are an achievement, to be worked at and valued.[53] In book v of *The Search After Truth*, on the sources of error in the passions, Malebranche takes up large parts of two chapters to discuss the ill and the good effects of wonder, of what happens 'when the brain is struck in places in which it has never been struck before, or when it is struck in an entirely new way' (LO 375, v.7). Wonder can work ill effects through the dangerous traces formed by violent and unruly animal spirits (chapter 9 below; Sutton 1998). But Malebranche does spend a little time acknowledging that, of the passions, only wonder 'illumines the mind', making it alone potentially 'useful to the sciences' (LO 385, v.8). This is because 'In wonder, the animal spirits are forced toward those parts of the brain representing the new object as it is in itself; there they make distinct traces that are deep enough to be preserved a long time. Consequently, the mind has a sufficiently clear idea of the object and easily remembers it' (LO 385, v.8). Where other passions move the spirits so that 'they represent objects only according to their relation to us and not as they are in themselves', wonder seems, sometimes, to allow acontextual remembering.

It looks, then, as if Malebranche thinks that clear and distinct remembering requires the difficult isolation or localising of each memory trace from others. Wonder might be the limiting case in which this happens, when one trace is distinct, deep, and independent enough to be preserved explicitly for a long time.[54] This contrasts with the normal case on a distributed model, which is the superposition of traces in one system whereby each particular trace is not itself explicitly preserved a long time. It is very hard, Malebranche says, to apply oneself to something which fails to excite wonder, 'since then the animal spirits are not so easily led into those parts of the brain necessary to represent it' (LO 385, v.8). This is a sign of some wish or hope for local representation, a desire that distributed memory not be all that we have.

It is worth just mentioning again that, beyond the general institutional

53 Medieval memory education, Mary Carruthers argues (1990: 7–12 and *passim*), had been a moral education, aimed at avoiding the murky forests of confused memories by institutionalising to-be-remembered material in rigid random-access cognitive formats, thus encouraging prudence, piety, and good citizenship. The source material which Carruthers analyses primarily in terms only of spatial metaphors for memory reveals the differing moral implications of local and distributed models of memory. I address wider implications of these moral memory systems in Sutton 1997.

54 Explicit representation seems to guarantee immunity from melding. This is clear on a generally localist model, which is set up to exclude interference (see on Hooke, chapter 5 below). But it is also true of a predominantly distributed model: when (very rarely, as in wonder) a single pattern of activation is explicit over the parts of a physical system in which other traces are also stored, this means that there have been insufficient changes to the patterns and interrelations of the physical parts which store all these traces to cause the explicit representation's degeneration, transformation, or condensation.

strength of Cartesian natural philosophy in general (Heilbron 1982: 26–38) and physiology in particular (Sloan 1977), the animal spirits theory of memory gained immense popularity over the late seventeenth and early eighteenth centuries. From Descartes until, roughly, the 1740s, remembering was the motion of animal spirits through the pores of the brain (Yolton 1984b: ch. 8). The rife associationism of the eighteenth century, in fields from aesthetics and ethics to psychology and physiology, derived much of its plausibility from such earlier distributed models. Even if Hume was right to identify a neurophilosophy as specifically Cartesian, it was soon taken up by writers of differing metaphysical persuasions. Malebranche's strong affirmation of the centrality of the spirits for psychological explanation set the tone: 'all the changes occurring in the imagination and the mind are only the consequences of those encountered in the animal spirits and the [brain] fibres' (LO 134, II.11.2). The animal spirits had the entire realm of mental processes in their wriggling power.