Consistent with the spirit of Locke and Lashley, current connectionist models may argue for a memory as not existing locally, and as being realised only on retrieval.

(Herbert F. Crovitz 1990: 174)

DIDEROT: Could you tell me what the existence of a sentient being means to that being himself?
D'ALEMBERT: Consciousness of having been himself from the first instant he reflected until the present moment.
DIDEROT: But what is this consciousness founded on?
D'ALEMBERT: The memory of his own actions.
DIDEROT: And without that memory?
D'ALEMBERT: Without that memory there would be no 'he', because, if he only felt his existence at the moment of receiving an impression, he would have no connected story of his life. His life would be a broken sequence of isolated sensations.
DIDEROT: All right. Now what is memory? Where does that come from?
D'ALEMBERT: From something organic which waxes and wanes, and sometimes disappears altogether.

(Denis Diderot, 'D'Alembert's Dream' (1769/1964: 155–6))

Introduction

Responses to Descartes' distributed model revealed perceived connections between theories of memory and wider views about human nature: how ordered or chaotic were cognitive processes thought or desired to be? This chapter explores more explicit relations between the animal spirits model of memory and concerns about psychological unity and order, through philosophical discussions of the continuity of personal identity. How likely were the fleeting spirits to preserve sameness of personhood over time?

For John Locke, a hierarchy of dependence ran from religion and morality through personal identity, by way of consciousness as extended by memory, which rested in turn on fleeting animal spirits. The vulnerability of his scheme became increasingly obvious. Locke's critics complained that he reduces a person to a club of jostling spirits in the brain: since true memory is agreed by
all to be impossible in fleeting matter, worried Samuel Clarke, Locke's anchoring of self in memory means that we are all 'unavoidably we know not who, and do but fancy and dream ourselves to be the Persons we think we are' (in Fox 1988: 54–5, 144–5).

In chapter 9 I further adumbrate the perceived immorality of the animal spirits by looking at their unwholesome associations with contagions of the imagination, seductive and garish images, demonic action, and male sexual insecurity. Conservative critics of Locke questioned the reductive flow from morality through memory to inconstant spirits at the first hurdle, by denying the link between the thinking substance of the self and any psychological (let alone physiological) process. But Locke's refusal to rest the great questions of responsibility and accountability for action on obscure theological ontology was influential. Immaterial substance no longer being certain, the person might be but 'a system of floating ideas'. In the Lockean world of the eighteenth century, that hierarchy of dependence was harder to resist, and 'moral Man' would be saved only by the introduction of greater stability and continuity into physiology by eliminating the animal spirits (chapter 10).

7.1 Memory and personhood: a physiological puzzle

The puzzle

Theories of distributed memory afford an unusual perspective on Locke's views of personal identity. Even if we do not accept G.S. Rousseau's claim (1969/1991: 4) that Locke's 'deepest questions are ultimately physiological', neurophilosophical themes are at work in more of his discussions of psychological phenomena than commentators generally admit. It is possible, in particular, to come to grips with a strange physiological puzzle in Locke's account of the self, the first major treatment of the modern 'problem' of personal identity.

Locke's new theory of personal identity, summarised with the maxim 'consciousness alone makes self' (Essay II.27, heading to 23–5), fuelled immediate controversy (Fox 1988) and is still often taken as a basis for the construction of philosophical theories of self (Shoemaker 1963; Wiggins 1976; Parfit 1984: 205ff.; Wilkes 1988a). Yet there is a neglected oddity towards the end of the long chapter on personal identity added to the second edition (1694), when Locke reflects on 'some suppositions that will look strange to some readers' (Essay II.27.27). In line with his general hostility to essentialist views of identity which located selves in non-physical souls (Allison 1966/1977), Locke remarks again on our 'ignorance of that thinking thing, that is in us, and which we look on as our selves'. He continues with this worry:

Did we know what it [that thinking thing] was, or how it was tied to a certain System of fleeting Animal Spirits; or whether it could, or could not perform its Operations of Thinking and Memory out of a Body organized as ours is; and
whether it has pleased God, that no one such Spirit shall ever be united to any but one such body, upon the right Constitution of whose Organs its Memory should depend, we might see the Absurdity of some of these Suppositions I have made. (Essay II.27.27)

I do not know of any extended discussion of this passage by commentators: but this is a strange, strong disclaimer. If we knew more about ‘a certain System of fleeting Animal Spirits’ and about the right constitution of the bodily organs on which memory depends, Locke’s new account of the person might look absurd! Why this tension?

It is hard to locate the source of Locke’s worry in this paragraph. One concern, as earlier in the chapter (II.27.13–15), is with ‘strange Suppositions’ about transmigration: can God’s goodness be relied on to prevent the existence of many persons in one body or the same person in many bodies, for which Locke’s theory seems to allow (Curley 1982: 305–6, 310–14)? Since the ‘Spirit’ about which Locke wonders must be immaterial substance, he is also reiterating doubts about the utility of reference to such thinking things in this context, on the grounds that it will be impossible to individuate persons by reference to something of which we are ‘in the dark’ (11.27.27). But why are these more familiar points raised here in the context of possible connections between thinking substance, animal spirits, and the bodily organs of memory?

1 Along with most other references to physiological psychology, it is omitted from A.D. Woolley’s abridgement of the Essay (chapter 1 above; on abridgements compare Alexander 1985: 2–3). Yolton (1984b: 158–9) notes the mentions of animal spirits here and at II.27.13 (see below), without examining the suggested threat to Locke’s suppositions. He argues rightly that Locke resists the identification of thought with matter which others would ascribe to him. But Locke’s language here is looser than that of identity: thought could be ‘tied to’ matter in many ways without being matter. I suggest that only material constraints on cognition are necessary to raise trouble for Locke here, and that Locke accepted that there were such constraints.

2 Immortality is the central context for all early modern debates about personal continuity. Responsibility at time of judgement seemed to require a strong form of identity between the sinning agent and the judged sinner. For the complexity of orthodox views on the resurrection of the body see Davis 1988; Bynum 1992, 1995b.

3 Michael Ayers, to whom I am extremely grateful for helpful conversation and correspondence, disagrees with my reading of the puzzle. Ayers takes the ‘suppositions’ which conflict with physiology to be not Locke’s own views about personal identity, but the set of thought experiments used to demonstrate the irrelevance of immaterial substance to problems of individuation and accountability. The reference of ‘suppositions’ is not clear. Certainly, Locke’s readers did not need to refer to a particular physiological theory to criticise the structure of his views, which seemed wrong-headed in basing personal identity in a ‘mode’ like consciousness, since modes are by nature fleeting, rather than in immaterial substance. But my suggestion is that both Locke and some critics did at times recognise specific problems arising, for his own views, from those theories of memory and its physiology which did not allow items to be stored and recalled independently one from another. Ayers accepts the general threat posed by animal spirits accounts of memory to personal identity. His own clear account of Locke’s theory and its contemporary critics is in Ayers 1988: vol.11, 260–77.
I approach this question in roundabout fashion, looking at its historical and philosophical aspects. I show how Locke’s treatment of memory responds to theoretical issues about local and distributed representation, and for this reason plays an ambiguous role in his account of the person. Firstly, though, I need to set out briefly the relevant aspects of the influential chapter on identity.

**Memory, consciousness, and self**

Though the details are hard to work out, memory plays some central role in Locke’s account of personal identity. For Locke, consciousness, both in the present and as it extends backwards in time, is the sole criterion for sameness of personal identity. Personal identity thus depends not only on ‘that consciousness which is inseparable from thinking and, as it seems to me, essential to it’ (II.27.9) but also on memory.

For since consciousness always accompanies thinking, and it is that that makes everyone to be what he calls Self, and thereby distinguishes himself from all other thinking beings: in this alone consists Personal Identity, i.e. the sameness of a rational being. And as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that Person: it is the same Self now it was then, and it is by the same Self with this present one that now reflects on it, that that action was done. (II.27.9)

There is much dispute on the extent to which Locke makes continuity of self dependent on memory. Although eighteenth-century reactions to his theory assumed that Locke identified consciousness with memory, or made memory a sufficient condition for identity over time, this is by no means an inevitable interpretation (see the appendix to this chapter). It is simply that memory is, for Locke, at least one important form of the necessary kind of psychological continuity.

Locke uses his psychological criteria for personal identity to buttress the social functions played by the concept of a person. There are thus two strands in Locke’s treatment, one ontological and psychological, the other normative and ‘forensic’. The importance of the latter strand lies in its links to moral, legal, and social institutions and practices. The relation between the two strands is not entirely clear in Locke, and he has been criticised for confusion between them.4 Tensions between the ontology of the self and

---

4 Tennant (1982: 74) laments Locke’s ‘linguistic errors in bringing together the ontological and the forensic’ and the way in which he ‘inextricably links and confuses the two issues’. Mary Douglas (1992: 46–9) argues that Locke’s new vision is driven only by ‘fit with legal and economic institutions’, implausibly grounded in theological necessity: ‘a unitary, responsible self-agent must be supposed to exist because it is intellectually, juridically and morally necessary’ (1992: 49). See also Mackie 1976: 183. Ross Poole and Catriona Mackenzie tell me that criticism of Locke for confusion on this point has been a common theme in discussions of personhood in applied ethics and social philosophy, especially when critics take the social strand to be in fact primary and consider the ontological and psychological strand dispensable. Compare Poole 1992, 1996.
consideration of what Locke called the 'moral Man' who is a 'corporeal rational Creature' (III.II.16; Behan 1979: 61–3) fuel the remainder of this chapter.

'Moral Man'
What are the elements of this strong conception of self? In looking at responses to Descartes in chapter 5 I showed the tension between distributed models of memory and certain strong conceptions of psychological order and control. Through Locke, tensions with interpersonal and with social/institutional requirements can be added. ‘Person’, for Locke, is ‘a Forensick Term, appropriating Actions and their Merit; and so belongs only to intelligent Agents capable of a Law, and Happiness and Misery’ (II.27.26).

This normative concept is the basis of moral responsibility, and is the ultimate source of ‘all the Right and Justice of Reward and Punishment’ (II.27.18; Allison 1966/1977: 109). A person will ‘extend itself beyond present Existence to what is past’ through concerned consciousness, consciousness ‘founded in a concern for Happiness’ (II.27.26). Only in appropriating past actions as one's own, or by reconciling a present self to them, are persons accountable. The juridical system requires such accountability, for punishment is ‘annexed to personality, and personality to consciousness’ (II.27.22). Rationality is a further condition for these ethical, legal, and social norms: personal identity, ‘the sameness of a rational being’ holds only for ‘a thinking intelligent Being, that has reason and reflection’ (II.27.9). The absence of either rationality or concern for actions is taken, in Locke's thought experiments, to be just as strong evidence against a being's personhood as is total amnesia. The forensic nature of the concept of person implies that not all humans will be persons. Some will be excluded as not meeting the required standards which only ‘intelligent Agents capable of a Law’ meet.5

This idealised concept of person, then, must bear a lot of weight. It is persons who are truly free rational agents, who are the loci of moral responsibility, who are the individual units in a just political society. And, although consciousness can also extend forwards, memory is one important means for the extension backwards of concerned consciousness to past actions or thoughts.

5 Locke takes the gap between human and person to be a strength of his theory: while we have to ascribe responsibility and agency to a 'rational Parrot' (II.27.8), we need a separate set of (biological) criteria for the identity of a human (II.27.1–7) which do not presuppose intelligence and capacity for a law. The talking parrot was denounced as 'devilry' by Richard Burthogge (1694/1976: 13–22).
7.2 The spirits and the soul
The ‘problem’ of the self
It is far from obvious on ‘internal’ grounds alone why the concept of person should have come to concern Locke relatively late in his philosophical thinking. The philosophical ‘problem’ of personal identity, in the form of a search for conditions or criteria for identity of self across time, was not a traditional one, but came to exercise philosophers rather suddenly: Hume, depressed by his own failure to catch his ‘self’, reflects within fifty years of Locke’s Essay that the question of personal identity ‘has become so great a question in philosophy, especially of late years in England, where all the abstruser sciences are study’d with a peculiar ardour and application’ (Treatise i.iv. 6: 259).

The causes of the peculiarly English concern with personhood from the late seventeenth century are obscure and various. Relevant factors include dissatisfaction with Cartesian and theological essentialisms, new concern to justify the location of basic political and economic rights and responsibilities in the individual, the imposition of order and unity in religious contexts, and a parallel concern in political ideology with affirming national identity through national history. If sameness of nationhood over time is guaranteed and extended backwards by historical continuities, then memory will function to provide analogous access to continuities in the individual past. Changing usage in the seventeenth century of words like ‘individual’ (Williams 1976: 162–3) and ‘consciousness’ (Wilkes 1984, 1988b; Hagstrum 1987; Thiel 1991), as well as ‘person’, mark the conceptual shifts, although recent cautions about the arbitrariness and futility of generalised searches for origins in the history of the self (Tomaselli 1984a; Skinner 1991; Reiss 1996) warn us against taking etymology as more than suggestion. Without here trying to unravel these larger changes, I want to approach the historical aspect of Locke’s physiological puzzle. Why, firstly, would anyone think that the self or soul was ‘tied to a certain System of fleeting Animal Spirits’?

Her invisible self
It had been common in ancient and Renaissance physiology to see the spirits as the immediate instruments of the soul (chapter 2 above). But in the seventeenth

6 As Fox argues (1988: 28, 140), concerns with Descartes’ cogito and problems about thinking substance, which had worried Locke at least since 1682, were part of the motivation. I discuss below one earlier remark of Locke’s on personhood and corporeal spirits.

7 So Hume (Treatise i.iv.6, p. 261) compares the soul to a commonwealth, remaining identical through total changes of its laws and members; this is evidence that memory produces, rather than discovers, (the fiction of) personal identity, as history produces national identity. On the production of group history as national memory by the English, constructing themselves as the new Protestant elect and Anglicanism as the true faith lost behind the corruptions of religious history, see Trevor-Roper 1989: 120–65; compare Pocock 1971; Rattansi 1988.
century, some began to think specifically that the animal spirits might provide some privileged kind of access, unavailable in any other way, to the innermost interiority of the soul. The topic is addressed in detail by Henry More. Though hostile to Cartesian uses of animal spirits in distributed models of memory (chapter 5 above), More argues, in defending *The Immortality of the Soul*, that corporeal animal spirits are ‘the immediate engine of the Soul in all her operations’ (IS II.8.3: 95; on More’s version of animal spirits theory, see Iliffe 1995: 437–9). Because the spirits are found ‘in their greatest purity and plenty’ in the fourth ventricle, this is where ‘that precious and choice part of the Soul which we call the Centre of Perception is to be placed’: the fourth ventricle, whence the soul can command the spirits, is suitably called ‘the Root of the Soul’ or ‘the Eye of the Soul’ (IS II.8.2: 95; II.11.10: 109).8 The intimate connection between spirits and soul explains the instantaneous effects of the soul’s commands, for the spirits’ ‘Swiftness of Motion is much like that of Light, which is a Body as well as they’ (IS II.9.4: 100).9 The soul, which cannot move matter but only determine it in motion, is in immediate contact with the spirits as they are ‘playing about and hitting against the sides of the Caverns they are in’, and in voluntary motion ‘she, when they are playing onely and gently toying amongst themselves, sends [them] forth into the exteriour members’ (IS II.8.10: 97; II.9.3 100). Arguing against those who believe that something other than animal spirit is the immediate seat of the soul, More tries metaphorically to ward off any danger arising from the spirits. But he has to accept their potentially harmful effects on the soul:

And it is no wonder, if the continuation and natural composure of the Spirits be Rest and Ease to the Soul, that a violent disjoyning and bruising of them, and baring the Soul of them, as I may so speak, should cause a very harsh and torturous sense in the Centre of Perception. (IS II.10.8: 104)

More, then, advises us to ‘contain our selves within the capacities of the Spirits’: the frequency, vividity, or novelty of an impression can, through the action of these spirits, ‘pierce the Soul’, sometimes ‘with an extraordinary resentment’ (IS II.11.5: 107). He hopes to support a substantialist view of the self, by showing the ontological gulf between the corporeal spirits and that to which they had access. It is important, he reminds us, ‘that the spirits are not

---

8 The only important function of the entire brain is to act, with the nerves, as ‘one continued Receptacle or Case of that immediate Instrument of the sensiferous motions of the Soul, the Animal Spirits’, and thus protecting the ‘chief seat and Acropolis of the Soul’ in the fourth ventricle (IS II.8.13: 98).

9 This is to answer the ‘frivolous objection’ of ‘simple and careless . . . Opposers of this ancient and solid Opinion’, that corporeal motion could not be quick enough (IS II.9.4: 100, and II.9.1: 99). More does not think it ‘so hard a business that these Spirits should be commanded downwards into the Nerves’. The effects of different atmospheric conditions on the spirits too explains why our thoughts are clear in clear air ‘and in cloudy more obscure and dull’ (II.8.4: 95).
sufficient of themselves for these Functions; nor the Soul of her self, without
the assistance of the Spirits' (IS II.11.1: 106). But, as Henry (1986b) argues,
emphasising the containment of the self within the capacities of the spirits
could also lead towards materialism: if the spirits can do so much of the soul's
work, the need for reference to the soul could drop out. One possibility, then, is
that since Locke had little sympathy either for substantialist views of self or for
materialism, he could not be expected to approve of talk of the spirits' access to
an invisible soul.

The idea of special intimacy between soul and animal spirits continued to
fascinate. In a dialogue on the 'Volatile Oeconomy of the Brain' which
synthesises early eighteenth-century theory, Mandeville's mouthpiece Philopirio praises

the transcendent subtily of those airy velocious Agents, the chief and
immediate Ministers of Thought; that officiating between the Soul and the
grosser Spirits of the Senses have always access to her invisible self.
(Mandeville 1711/1976: 131)

Strange consequences follow from this intimacy, for the spirits are inti-
mately connected not only to the soul but also to grosser functions from which
moral philosophers officially wanted to keep the self clear (chapter 9 below).
Locke’s new psychological criterion for sameness of personhood combines
with the peculiarly modern attempt to penetrate the recesses of the (female)
soul, inviting heroic quests into the mysterious interior (Tomaselli 1984a:
198–201; compare Dollimore 1984/1989: chs. 1, 5, 10, 16; Barker 1984; Belsey
1985: ch. 2; Sutton 1990). But to my knowledge the link between his theory of
personal identity and the seventeenth century's 'shadow of physiology' in the
form of the animal spirits has not yet been noticed.

Sydenham on self
Firstly, it is worth pointing out who Locke is answering. The phrase used by
Locke in the passage from the Essay which set our puzzle was 'a certain System
of fleeting Animal Spirits'. The phrase echoes the following intriguing passage
from a 1682 discussion of hysteria by Locke's medical friend, mentor, and col-
league Thomas Sydenham:

Beyond what we may call the exterior man, who is composed of parts which
are visible to the senses, there is an interior man formed of a system of animal

10 Also quoted by Monro 1975: 62. Mandeville changes his name for Philipirio (1711/1976:
xi). Despite official caution about hypotheses, Mandeville/Philopirio devotes long tracts of
this second dialogue (1711/1976: 121–48) to speculation on internal mixtures, to give the
hypochondriacal Misomedon 'a clear Idea of the broken contexture of [his] Spirits'
11 The phrase, for those intrusions of physiology which irk the dualist, is Keith Campbell's
spirits, a man who can be seen only with the eyes of the mind. This latter man, closely joined and so to speak united with the corporeal constitution, is more or less deranged from his state to the degree that the principles which form the machine have a natural firmness. That is why this disease attacks women more than men, because they have a more delicate, less firm constitution, because they lead a softer life, and because they are accustomed to the luxuries and commodities of life and not to suffering.

Sydenham's links with Locke have been well documented (Niebyl 1973: 369-72; Romanell 1984: ch. 4; Sanchez-Gonzalez 1990). Locke's papers often mention Sydenham's reliance on animal spirits in therapeutic practice. In 1679 he records that Sydenham informed a Mrs Duke that her false conceptions proceeded 'from want of spirits and coldness in the habit and womb' (in Dewhurst 1963: 177). In 1680 Sydenham told Locke that he had cured the Earl of Salisbury of a 'total suppression of urine' which was 'from a disorder of the spirits' (in Dewhurst 1963: 193). These examples remind us of a practical dimension of seventeenth-century medicine in which spirits, humours, and fluids continued to drive treatment. The theoretical role of spirits made this inevitable.

Both mental and physical symptoms of hysteria 'proceed from a Confusion...'
[ataxia] of the Spirits' in various kinds of 'unequal Distribution, which is altogether contrary to the Oeconomy of Nature'. Ataxy and disorder in the spirits occur when 'too many of them in a Croud, contrary to proportion, are hurried violently upon this or that part', perverting organ functions: this explains too hysteria's characteristic protean adaptability to the peculiar symptoms of different body parts (Sydenham 1682/1685: 132–3). An offence to corporeal harmony, the contrary spirits seem by nature compelled to interfere, thwart, and obstruct proper order in memory and body (Glanvill, VOD: 39; chapter 5 above).

I return in chapter 9 to the threat to sanity posed by the animal spirits theory of memory. But already there is reason to think that the puzzling passage in Locke's essay is aimed in part at the way Sydenham tied an 'interior man' to a system of animal spirits. It was, indeed, in 1683 that Locke first articulated his unusual view of sameness of personhood over time:

Identity of persons lies not in having the same numerickall body made up of the same particles, nor if the minde consist of corporeal spirits in their being the same, but in the memory and knowledg of ones past self and actions continued on under the consciousness of being the same person whereby every man ownes himself. 16

Even if matter, in the form of corporeal animal spirits, could think, only psychological continuities of memory and consciousness could guarantee sameness of personhood. Animal spirits, like all material particles, do not remain stable or constant over time, and so personal identity is to be sought not in physiological identities across time but in the extension backwards, through memory, of consciousness and in the appropriation of one's past self and actions. Yet by the time he writes Essay II.27.27, even after expanding these suggestions about psychological continuities and examining their implications for puzzle cases, Locke feels residual tension between the animal spirits psychophysiology and the important role he has given to memory in discovering sameness of consciousness over time. I suggest this is because he has also been thinking more about memory, and has found little to ground confidence in the kind of continuity which persons must have if the forensic functions of the concept are to be fulfilled.

7.3 Amnesia and identity: the fragility of memory
With some historical background to Locke's puzzle in place, I return to the philosophical issue. Why does Locke fear the collapse into 'absurdity' of his account of self if memory depends on 'the right Constitution' of certain bodily

16 Locke's journal, Tuesday 5 June 1683 (British Library MS Locke, fo. 7, p. 107), in Dewhurst 1963: 222. I am grateful to Udo Thiel for bringing this note to my attention.
organs and ‘a certain System of fleeting Animal Spirits’? The answer is in the
detail of his views on the psychophysiology of memory.

Locke on memory
Locke is officially unwilling to ‘meddle with the physical consideration of the
mind’ (Essay 1.1.2). But he does assume his readers’ familiarity with the ubiqui-
tous late seventeenth-century animal spirits physiology (compare Wieand
1980: 69–70). He is as sympathetic to it as to the other ideas in natural philoso-
phy which were being put forward by such as Boyle, Sydenham, Huygens, and
Newton, the four ‘master-builders’ of the commonwealth of learning for
whom Locke claims to be a philosophical under-labourer (Essay, ‘The Epistle to
the Reader’, pp. 9–10).17

In the chapter on retention, Locke is reluctant to enquire into the extent to
which memory might depend on ‘the Constitution of our Bodies’, ‘the Temper
of the Brain’, and ‘the make of our animal Spirits’ (Essay II.10.5).18 But he
admits that it seems ‘probable’ that ‘the Constitution of the Body does some-
times influence the Memory’. He cites examples of how diseases and fevers can
‘calcine all those Images to dust and confusion, which seem’d to be as lasting,
as if graved in Marble’ (Essay II.10.5). Ordinary remembering and forgetting, he
accepts, are probably influenced by multiple corporeal factors even in health.

In the first edition of the Essay Locke writes that ‘Memory ... is as it were the
Store-house of our Ideas’, the ‘Repository’ where ideas are laid up (Essay
II.10.2). This seems a straightforward localist storehouse theory like Hooke’s,
by which ideas are statically ‘lodg’d in the Memory’ (Essay II.10.7). But in the
second edition, responding to criticism from Norris,19 Locke adds the claim, or
disclaimer, that

this laying up of our Ideas in the Repository of the Memory, signifies no more
but this, that the Mind has a Power, in many cases, to revive Perceptions, which

17 Locke’s library (Harrison and Laslett 1971) included works by many authors in the spirits
tradition, including Descartes, Digby, La Forge, Malebranche, More, and Willis. On
Locke’s uses of Willis’ physiology see Wright 1991a.
18 The phrase ‘and the make of our animal Spirits’ was not added until the fourth edition of
1700. But this does not mean that he did not accept their particular relevance earlier:
motions of animal spirits in birds, he notes in the first edition, can mechanically ‘leave
traces in their Brains’ (II.10.10). Other mentions of animal spirits (II.1.15, II.8.4, II.8.12,
and II.8.21) refer to their role in perception. Compare references at IV.10.19 (2nd edn) and
II.33.6 (4th edn), and Yolton 1993: 11–12. One earlier connection between memory and
corporeal spirits is in Locke’s journal for Thursday 1 July 1677, while in France: ‘for the
memory they take the dried flowers of sage and rosemary infused in the spirits of wine
which they use’ (Dewhurst 1963: 83). Compare Hooke’s more grating schemes for
improving memory by the intake of various metals (chapter 5 above).
19 Norris pointed out that ideas could not, on Locke’s view, really be stored or ‘as it were laid
aside out of Sight’ (Essay II.10.2), because all ideas, for Locke, must be currently perceived
(Essay I.2.5). The relevant passage in Norris is quoted by Wieand (1980: 65–6). On Norris’
it has once had . . . in this Sense it is, that our Ideas are said to be in our Memories, when indeed, they are actually no where, but only there is an ability in the Mind, when it will, to revive them again. (II.10.2)

Stressing again his conviction that ideas, as ‘actual Perceptions in the Mind’, must ‘cease to be any thing, when there is no perception of them’ (II.10.2, added in the second edition), Locke says that it is only metaphorically that we talk of ideas being (stored) in our memories.

Critics of trace theories from Reid to Norman Malcolm (1977: 195–7) have taken Locke to be confused or unclear in his attitude to the storehouse metaphor or model over the two editions. Yet inconsistency disappears if Locke is seen as moving towards the Cartesian theory of memory. Ideas do not remain explicitly in memory over the period from experience to remembering. What is ‘stored’ is not the idea itself, as in Hooke’s localist model: in one sense nothing at all is stored,20 for all that happens in ‘laying down’ a memory is that a change occurs in the ‘temper of the brain’ after which a physical disposition exists by which an idea can in certain circumstances be ‘revived’ or reconstructed. In a distributed model of memory,

patterns which are not active do not exist anywhere. They can be re-created because the connection strengths between units have been changed appropriately, but each connection strength is involved in storing many patterns, so it is impossible to point to a particular place where the memory for a particular item is stored. (Hinton, McClelland, and Rumelhart 1986: 80)21

It is, then, plausible to identify a memory trace not with a specific explicit pattern of activity in the motions of spirits, but rather with an implicit disposition for the evocation of the explicit pattern. As Locke claimed, we say that our ideas are (dispositionally) in our memories, when indeed they are actually (explicitly) nowhere.22 This distinction between explicit and implicit repre-

Wieand (1980: 72–3) says that Locke must be understood as taking impressions of the senses to be stored, impressions ‘from which we are able to generate (if we remember correctly) an idea which is numerically distinct from, but in certain respects descriptively the same as, the idea originally generated by the impression’. But this does not seem quite right, since impressions are fleeting; only dispositions endure.

This feature of distributed representation is ‘the main difference from a conventional computer memory’: it is certainly not a storehouse in which sacks of grain lie passive and unchanging save for a little decay until hauled out again. Where Aaron (1955: 138) wonders if Locke has completely given up the repository theory, Wieand (1980: 67) remarks that ‘there is good reason to suppose that Locke never held the storehouse theory’.

It remains unclear what sort of similarity exists between the original memory and that which is revived. Distributed memory traces, with their tendencies towards interference, can easily lose their identity. David Krell (1990: 76–7) grants, on the basis of this passage, that for Locke ‘memory is more construction than conservation’, but goes on to question whether his voluntaristic assurance in the ability of the mind to revive ideas ‘when it will’ is justified by his own views on forgetting and decay. This parallels my strategy of eroding confidence in the comforts which memory holds for the self: see below on control.
sentations requires a mechanism for causal continuity between experience and remembering without simple storage: only a physiological basis for the future revival of a trace can remain over time.\textsuperscript{23}

Locke's references to powers for the re-evocation of memories imply a metaphysics of physical dispositions, patterns, or textures. The patterns can come originally from the external world: in a sentence of Draft B of the Essay which did not find its way into the final version, Locke says that one way 'of retaineing of Ideas'

\begin{quote}
is the power to revive again in our mindes those imaginations which the objects from whence they came first caused in us when they affected our senses whether by motion or otherwise it matters not here to consider. (Draft B, Locke 1990: sect. 23, p. 134)
\end{quote}

Peter Alexander (1985) has shown how the view that patterns can be transmitted from the 'textures' of objects to the 'textured' motions of animal spirits in the brain runs through the corpuscularian philosophy which Locke shared with Boyle.\textsuperscript{24} Different 'particular Textures of Matter' receive impulses differently (Burthogge 1694/1976: 158). These patterned motions are the basis for all remembering.

But this psychophysiology is in two kinds of tension with the need for memory to discover and to ground, in part, the continuity of personal identity over time. Even if memory can sometimes extend a self backwards in time to past consciousness and actions, it is threatened by both intrusions and interruptions. Firstly, the causes of present motions may not always be simply the original patterns from objects. This leads to problems in distinguishing memory from imagination and in controlling what is associated in memory. The intrusions of inappropriate ideas into the memory of 'moral Man' are my topic in chapter 9. The second threat to the Lockean self derives from the impossibility, on the model of implicit memory traces which he accepts, of the perfect preservation of motions and thus of memories. Locke laments that 'the ideas, as well as Children, of our Youth, often die before us' (Essay II.10.5). The defects of memory mean, then, that prospects are poor, on Locke's theory, for perfect continuity of personhood between youth and maturity.\textsuperscript{25}

\textsuperscript{23} Locke's belief in the physiology of animal spirits leads him to say that it is 'impossible' to refuse to grant memory to 'several other Animals'. The sounds birds hear make traces which 'by their after-endevours' allow the production of similar sounds (Essay II.10.10). The ascription of memory and perception to animals is not anti-Cartesian, as Wright (1991a: 255) claims: Descartes' beast-machines too dream, feel, and remember (chapter 3 above).

\textsuperscript{24} See especially Alexander's chapter 4 on ideas, chapter 7 on powers, and chapter 10 on patterns, with a discussion of the animal spirits theory of memory on pp. 192–3. Alexander's reading depends on taking Locke's corpuscularian scientific realism seriously. I cannot defend this here, but a classic statement of the case is Mandelbaum 1964.

\textsuperscript{25} This would push Locke's account of personal identity into a familiar circularity if memory was the only kind of psychological continuity which he allowed as a criterion. But if other
Locke on forgetting, physiology, and self

In paragraphs on forgetting and other failures of memory, Locke refers frequently to psychophysiological factors. This is more than religious degradation of the body and attribution of human failings to the corporeal: there is specific reason for memory's inability to reproduce the past in all its desired presence within the theory of memory.

Sections 4 and 5 of Essay II.10 describe how 'ideas fade in the memory'. At times Locke talks conventionally of how deep an imprint needs to be, of the mind's regular failure to set 'the stamp deep into itself' (Essay II.10.4). Mixing metaphors, he describes regretfully the wearing out of the print and the fading colours of 'the pictures drawn in our Minds', even of 'those which are struck deepest'. Ideas can vanish, even from 'Minds the most retentive', 'leaving no more footsteps of remaining Characters of themselves, than Shadows do flying over Fields of Corn' (Essay II.10.4–5). Even conscious attempts at careful remembering are vulnerable to weakness in memory 'either through the temper of the Body, or some other default' (II.10.4). Without the requisite repetition of recall which (as Descartes' L'Homme had suggested) would refresh the memory, the fading ideas of the memory will 'vanish and disappear'. It is at this point that Locke is forced back to physiology:

> How much the Constitution of our Bodies, and the make of our Animal Spirits, are concerned in this; and whether the Temper of the Brain make this difference . . . I shall not here enquire, though it may seem probable, that the Constitution of the Body does sometimes influence the Memory. (II.10.5)

Forgetting, then, falls naturally out of the psychophysiological theory, as patterns traced in the animal spirits fail to stick, fade, or disappear entirely.

Locke's reluctance to enquire further into differences made by body, brain, and spirits is not so surprising. Memory, according to Locke, is, for 'an intellectual Creature' like a man, 'of so great moment that, where it is wanting, all the rest of our faculties are in a great measure useless' (II.10.8). But if memory depends on the dispositions and motions of the fleeting animal spirits, and personal identity depends on memory, and morality depends on identity, then the purpose of the whole Essay, the preservation of the 'great Ends of Morality and Religion' (IV.3.6), may be thwarted at the outset by the rare and subtle corporeal particles. Ontological caution is a necessary, prudent defence.

That this is the right direction for interpreting Locke's puzzle is perhaps confirmed by another reference to the animal spirits earlier in the chapter on identity which shows that Locke is alive to the kind of threat they pose. Noticing (in a perplexing passage) that his theory requires God benevolently to have continuities, like the connections between intentions and actions, or emotional continuities, can discover sameness of personhood over time for the adult, then there can be genuine questions about what kind of history this adult has.
ensured that the same memories and consciousness belong, as a matter of fact, to only one human, Locke remarks: 'How far this may be an argument against those who would place thinking in a system of fleeting animal spirits, I leave to be considered' (II.27.13). Even though it is only a few pages further on that Locke complains about absurdity in his own account if knowledge of the fleeting spirits were more advanced, it seems here that it is the animal spirits physiology which might not survive a proof of its incompatibility with the new theory of the person. Any tension between animal spirits and the forensic concept of a person might in time encourage revision of the psychophysiology in order to retain the normative notion.

Hume on forgetting, physiology, and self

I quoted at the start of chapter 3 Hume's approval of 'the Cartesian philosophy of the brain' in explaining an odd case of amnesia (in Hume 1948: 502, also in Wright 1983: 84 and Yolton 1984b: 188). Descartes' neurophilosophy provides a framework for understanding how connections with the personal past could be severed: might it also lie behind Hume's criticisms of the way Locke makes personal identity dependent on memory, and his own failed attempts to catch the self? For Hume the formation of our (fictional) idea of a unified self depends upon associative 'easy transitions' from one past perception to another in acts of memory (Treatise I.iv.6: 262). The associative relations of resemblance and causation on which 'identity depends', and which associate 'the whole train of perceptions' in the imagination, are both mechanisms of memory. Memory is 'to be consider'd . . . as the source of personal identity'.

26 I am very unsure about the import of this whole passage at II.27.13. God, caring about the 'Happiness or Misery' of his sensible Creatures', will not, says Locke, 'by a fatal Error of theirs transfer from one to another that consciousness, which draws Reward and Punishment with it'. I understand neither why this would be an error made by the creatures, nor what exactly it is which Locke thinks might be an argument against those who place thinking in the system of fleeting animal spirits. Nothing I have read on the passage throws any light on these questions. Flew (1951: 164), Mackie (1976: 184–5), and Taylor (1989: 543–4) are all preoccupied with convicting Locke of circularity, and argue that there would not be any error here in the first place because the transfer of consciousness would by definition also transfer identity. But, as I have suggested, psychological continuity does not only extend into the past, and so there are other criteria than memory for sameness of personhood. Alston and Bennett (1988: 42) and Wedeking (1990: 177–9) take the error to be in the unfairness done to the substance, not to the person, who (which?) is punished. But no one discusses the use of these considerations against those who place thinking in the animal spirits. As Wedeking says (1990: 178), it may be that this passage of Locke's 'will forever defeat any attempt to read the section as a coherent argument or line of thought'. Michael Ayers (personal communication) thinks that those who place thinking in animal spirits are not materialists, but dualists for whom the soul is meant to use spirits as instrument (when 'transfer' of one soul to another physical system would seem possible).

27 I cannot here go into detail on the interpretation of this chapter or the remarks on personal identity in the 'appendix'. Good expositions are Flage 1990: ch. 7 and Pears 1990: chs. 8–9; for a survey of views, with original interpretations, see Fogelin 1985: ch. 8.
because it 'alone acquaints us with the continuance and extent of this success-
ion of perceptions' which we thus naturally tend to think of as really bound
together (I.iv.6, 259–61).

But although the fact is not explicitly mentioned at this point in Hume's
'accurate anatomy of human nature' (I.iv.6: 263), we have already been told
that associative memory is psychophysiological. While Hume is reluctant to
examine the causes of associative principles, he does have a view, 'specious and
plausible', on the subject:

'Twou'd have been easy to have made an imaginary dissection of the brain, and
have shewn, why upon our conception of any idea, the animal spirits run into
all the contiguous traces, and rouze up the other ideas, that are related to it.
(I.ii.5: 60)

But the unruly spirits are by nature unreliable: it is true that they

always excite the idea, when they run precisely into the proper traces, and
rummage that cell, which belongs to the idea. But as their motion is seldom
direct, and naturally turns a little to the one side or the other; for this reason
the animal spirits, falling into the contiguous traces, present other related
ideas, in lieu of that which the mind desir'd at first to survey. (I.ii.5: 61)

Because we are not always sensible of this change in the ideas presented by
the spirits, their erratic motions are 'the cause of many mistakes and sophisms
in philosophy' (I.ii.5: 61).28 The spirits escape conscious notice, exceeding the
will.

Memory and psychological control
I return to the special dangers of misassociation in chapter 9: but Hume's
pessimism about personal identity will not be allayed by looking for any con-
stancy in the spirits' role in the processes of memory and association. Yet so
far I have given little weight to the voluntaristic talk which both Locke and
Hume keep up in describing the relations between self and psychophysilog-

28 See the clear discussion of Hume's psychophysiology of error in Wright 1983: 68–9,
71–4. Wright demonstrates Hume's proximity to Malebranche on these issues, and
suggests (1983: 73) that understanding the ties which philosophers 'believed to exist
between the continuous motion of fluids in the brain and the dynamics of our mental
lives' will aid insight into Hume's view of the self as 'no more than a continuous series
of perceptions related in certain ways'. There is, however, considerable debate (which
I cannot go into here) over whether or not natural relations among ideas are for
Hume, as Wright says (1983: 74), 'really neural in character'. This reading relies,
plausibly in my view, on a physical reading of Hume's talk of dispositions and natural
transitions in his chapter (I.i.8) on 'the causes of belief' (e.g. Treatise p. 99). Wright
defends this fully in his chapter 5, especially pp. 214–19. See also Anderson 1966:
118–24 and ch. 13; Anderson 1976; Wright 1991b: 150. For different views see
and connectionism.
archical processes. Will the mental origin of activities of recollection not suffice to guarantee sufficient continuity over time for the normative conception of the person as a rational, free, responsible moral agent to remain in place despite the obvious empirical frailties of human memory? Locke’s view, after all, was that ‘the Mind has a Power’ to revive ideas ‘when it will’: and, even with physiology in focus, Hume says ‘that as the mind is endow’d with a power of exciting any idea it pleases; whenever it dispatches the spirits into that region of the brain, in which the idea is plac’d’ (I.ii.5: 60–1) the various consequences described above follow. In both cases it looks as if the mind is meant to be genuinely ‘active’ in directing and controlling brain processes. But, not surprisingly, the details of this control over the corporeal continually slip away in the texts. 29 The idea of dominating one’s own body is little more than a wish, always already enveloped in the stronger seductions of surrender.

Locke announces that, in remembering,

the Mind is oftentimes more than barely passive, the appearance of those dormant Pictures, depending sometimes on the Will. The Mind very often sets itself on work in search of some hidden Idea, and turns, as it were, the Eye of the Soul upon it. (Essay II.10.7)

It is in passages like these that Locke seems most vulnerable to attacks on the kind of homuncular scanning, interpretation, or ‘viewing’ of inner ideas which critics of trace theories often denounce as incoherent. But the real difficulty with this talk is not in the assumption of memory traces but in the postulation of an active subject behind the traces which reviews and manipulates them as they lie passive in a memory store. In these passages, which date back to the early 1670s (Locke 1990: 135), Locke’s view of memory does seem to be a localist, storehouse one. But he is already aware, too, of the other phenomena which a distributed model takes seriously, and which threaten to undermine the desired executive control:

sometimes too they [memory ideas] start up in our Minds of their own accord, and offer themselves to the Understanding; and very often are rouzed and tumbled out of their dark Cells, into open Day-light, by some turbulent and tempestuous Passion; our Affections bringing Ideas to our Memory, which had otherwise lain quiet and unregarded. (II.10.7; also in Draft B, Locke 1990: 135)

This passage, acknowledging the effects of emotion on memory and the potential for unbidden activity among memory traces (which are after all dispositions in the pores of the brain and the continually flowing patterns of

29 This paragraph and the next are indebted to Krell 1990: 77–80. But Krell’s intention is to complain that ‘active reminiscence is given short shrift in Locke’s account’ (1990: 79), for Krell’s own view of memory requires a truly active subject to account for ‘mnemic output’ (1990: 89).
motions of the animal spirits),\textsuperscript{30} anticipates the chapter on association of ideas which Locke added to the fourth edition of the Essay in 1700 (chapter 9 below). Together with forgetting and the other defects of ‘oblivion and slowness’ in memory, such tumbling and confusion among memories suggests that the putative self, which needs the assistance of memory to ‘proceed beyond present Objects’ (II.10.8), has a severely curtailed or shrunken domain of untainted influence.

Development of my intrigue at this combination, at the coupling of official stress on the possibility of ‘moral Man’s’ control over his own physiology with repeated examples of the intrusion and spillage of the corporeal into his most exalted powers and activities, requires a foray into other metaphorical fields in which the quick and nimble spirits rummaged. Having suggested the intimacy between the animal spirits which form memory traces and the self, I can, in chapter 9, broaden the scope to include applications of the spirits theory to other mental, physical, and social phenomena. But first I return to the problematic search for the troublesome spirits.

Appendix: memory and self in Essay II.27

Until recently, some commentators assumed that Locke held a simple ‘memory theory’ of personal identity: passages like II.27.9, quoted in the text above (p. 160), were thought to imply that remembering doing something was both necessary and sufficient for being the same person as the person who did it. This view however seemed open to a barrage of criticisms. Sergeant, Butler, and, influentially, Reid complained that memory could not constitute personal identity, since memory presupposes personal identity: Locke’s theory thus seems circular, for nothing counts as remembering unless sameness of person between past and present is already assumed.\textsuperscript{31} Reid’s charge (1785/1849: 352) that Locke has confounded consciousness with memory was supported by modern critics from Flew

\textsuperscript{30} Compare, from Locke’s posthumously published 1692 reply to Norris: ‘ideas may be real beings, though not substances; as motion is a real being, though not a substance; and it seems probable that, in us, ideas depend on, and are some way or other the effect of motion; since they are so fleeting . . .’ (quoted by Winkler 1991: 218). The word ‘fleeting’ was associated for Locke both with ‘ideas’ and with ‘animal spirits’.

\textsuperscript{31} For these criticisms see especially Behan 1979. Locke’s own reply, taken from marginal notes in his copy of John Sergeant’s Solid Philosophy Asserted . . . (1697), is discussed by Thiel (1981). Both Behan and Thiel plausibly deny the force of the circularity objection, on the grounds that, for Locke, only the (natural) human is presupposed by remembering, not the (forensic concept of) person. Parfit (1984: 219–23) amends Locke by introducing a kind of memory, quasi-memory, defined so as not to presuppose the identity of the person who remembers, for example, having an experience with the person who had the experience.
(1951/1968) on. The notion, especially, that remembering having done something should be sufficient for having done it seemed, as Reid said, to conflate personal identity with our evidence for sameness of personhood: how could my suddenly remembering having done something suddenly make it true that I did it?

More recent scholars, then, querying the Reid/Flew strategy of interpreting Locke as exclusively a ‘memory theorist’ only to attack ‘memory theory’, have denied that Locke identifies consciousness with memory. The continuity provided by memory need not be the only kind of psychological continuity which a Lockean relies on: the same consciousness, indeed, ‘can extend to Actions past or to come’ (Essay II.27.10, my emphasis), so that continuity between, for instance, intentions and future actions too can be seen as important for sameness of personhood (and in turn actual disruptions to the links between action and intention will threaten continuity). A ‘functional distinction’ can be accepted between memory and the relevant kind of consciousness (Behan 1979: 66): for Edwin McCann (1987), ‘memory has its special role to play in personal identity only because of its connections with sameness of consciousness’, where ‘sameness of consciousness is the basic relation making for personal identity’. This explains why Locke uses the word ‘consciousness’ throughout the chapter while only referring explicitly to memory on a few occasions: yet memory is still left with a central role (Wiggins 1976: 150–1; Parfit 1984: 205; Schechtman 1994: 4–7). Memory is our main access to the past: for Locke, ‘memory is necessary if a person is to have a history’ (Behan 1979: 66). The ability to remember, in the present, having been an agent or having had certain experiences in the past is still something like a necessary condition, even if not a sufficient one, for being now the same person as the past person who acted or experienced.

The strongest point in favour of retaining memory as one central kind of psychological continuity even if it is not equivalent to consciousness is that Locke explicitly says that total amnesia rules out continuing personal identity. A being, whether material or spiritual, which has been ‘wholly stripp’d of all
the consciousness of its past Existence’, when such consciousness is lost ‘beyond the power of ever retrieving again’, may retain some other form of identity, such as identity of substance or identity of natural hu/man, but is not the same person as the earlier being (II.27.14; cf. II.27.20).\footnote{Atherton (1983: 284–5), denying the importance of memory, has to interpret this in a very forced fashion. She argues that, ‘when Locke is talking about amnesia’, he is thinking not of things which you thought or did but cannot now remember, but of ‘things you never thought or did and hence cannot remember’ and which are not ‘a part of you’ because you could never have remembered them. There seems no independent reason to believe this, and the way in which Locke sets up his second description of the case (II.27.20), specifically asking us to ‘suppose I wholly lose the memory of some parts of my Life, beyond a possibility of retrieving them’, looks \textit{prima facie} to confirm that the relevant actions or experiences have actually been forgotten (have once been available to memory, but are no longer).}