Locke's Philosophy of Mind

Jonathan Bennett

from: V. Chappell (ed), *The Cambridge Companion to Locke* (Cambridge University Press, 1994), pp. 89–114.

This chapter of the *Companion* will discuss the following topics in sections with these numbers: **(1)** Locke's acceptance of Descartes's view that there is a radical separation, a perhaps unbridgeable gap, between the world's mental and its physical aspects. Locke's view of **(2)** the cognitive aspects and **(3)** the conative aspects of the mind. **(4)** What Locke said about the possibility that 'matter thinks', i.e. that the things that take up space are also the ones that have mental states. **(5)** The question of whether all thought could be entirely caused by changes in the physical world. **(6,7)** What it is for a single mind to last through time. **(8)** What it is for a mind to exist at a time when it is not doing anything.

1. Property dualism

Descartes held a position that is sometimes called 'property dualism'. According to it, the properties that things can have fall into two classes—those pertaining to materiality and those pertaining to mentality—with no overlap between them. This is best understood as involving also a dualism also of concepts: the concepts that can be applied to things fall into two classes, with no concept in either class being reducible to or explainable through any belonging to the other class.

This property dualism can be felt all through Locke's *Essay*. He does not announce it as a thesis, any more than Descartes does, apparently accepting it as an unchallenged and unexamined axiom. While using facts about bodily behavior as evidence for conclusions about states of mind, Locke never asks *why* they are evidence (the 'other minds' problem seems to have begun with Berkeley); nor does he ever suggest that any cognitive concept might be analyzable in terms of behavioural dispositions or that sensations or feelings or 'ideas' might be physiological states.

Locke also accepts Descartes's view that minds must be transparent to themselves, for example in his polemic against innately possessed ideas and knowledge, where he says that we aren't aware of any such possessions and couldn't have them without being aware of them: 'To imprint anything on the mind without the mind's perceiving it seems to me hardly intelligible' (*Essay* I.ii.5: 49; see also II.i.11). But unlike Descartes he does not use this to define the realm of the mental, and it is not clear that he defines it at all. If he does, it is by saying that the idea of 'spirit'—which is one of his words for 'thing that has mentalistic properties'—is 'the idea of thinking, and moving a body' (II.xxiii.15). The second of those may seem odd: cannot bodies also move bodies?

Not really, Locke, thinks, because:

> When by impulse [a billiard ball] sets another ball in motion that lay in its way, it only communicates the motion it had received from another, and loses in itself so much as the other received. . . [This] reaches not the production of the action, but the continuation of the passion. . . The idea of the beginning of motion we have only from reflection on what passes in ourselves, where we find by experience, that barely by willing it, barely by a thought of the mind, we can move the parts of our bodies, which were before at rest. (II.xxi.4; but see the conflicting story in II.vii.8)

Unlike Descartes and his followers, Locke held no views about causation that posed any special problem for the idea of causal interaction between the material and mental realms, despite the categorial difference between the two kinds of property. We shall see that he allows not only that minds act upon bodies but also that bodies act upon minds.

The link between 'spirit' and 'mental' on the one hand and 'thinking' on the other does not help us much to grasp Locke's concept of mentality, because he gives no systematic account of what thinking is. In this respect, he does no better than Descartes though also, to be fair, no worse.

2. Cognition

'Thinking' and 'moving a body'—Locke's focus on these two fits with his statement elsewhere that 'The two great and principal actions of the mind. . . are these two: Perception or Thinking, and Volition or Willing' (II.vi.2; see also xxi.5,6). Locke's use of 'perception', and especially his relating of perceiving to having ideas, is chaotic. In one place, for example, he says that ideas are 'actual perceptions in the mind, which cease to be anything when there is no perception of them' (II.x.2). Nor does he say, carefully and consistently, what he means by 'thinking'. Still, in those formulations we can see him as expressing the view—held by many before and since—that mental doings fall into two large categories, the cognitive and the conative, or the intellectual and the volitional. This has been accepted and given a structural role by many recent philosophers who have sought to base a theory of mentality on the concepts of belief and desire.

At a quick glance, one would say that this leaves out two large mental matters: **(i)** emotions, feelings, passions, and **(ii)** sensory states, sense-data, qualia, phenomenal states, or the like. The nearest Locke gets to a treatment of **(i)** is in II.xx, 'Of Modes of Pleasure and Pain', in which he says that 'pleasure and pain. . . are the hinges on which our passions turn'. This chapter has its interest, but it doesn't contribute much to our picture of Locke's picture of the mind; and I shall not discuss it. As for **(ii)**: These appear in Locke's work as the having of 'ideas', which are treated in another chapter in this *Companion* and can be dealt with quickly here. The main point here is that Locke uses the term 'idea' not only for these sensory items but also for intellectual items that might be called 'thoughts' or 'concepts', these being the ingredients out of which beliefs are made. This is not an ambiguity in Locke's use of 'idea'; rather, he holds as a matter of theory that the mental items that come into the mind, raw, in sense perception are—after a certain kind of processing—the very items that constitute the basic materials of thinking, believing, and the like.

Setting aside, then, emotions and sensory states, we are left with the intellectual and volitional aspects of the mind, highlighted by Locke and also by a dominant trend in the recent philosophy of mind, namely the tendency to think that a proper understanding of mentality should be based largely on *belief* and *desire*. Let us see how these figure in the *Essay*.

To believe something is to believe *that P* for some propositional value of P. Locke's account of the rudiments of thinking is conducted in terms of 'ideas' (considered in their intellectual rather than their sensory role), and he takes these to be sub-propositional: he speaks of the idea of *horse*, of *man*, of *whiteness* and so on. In his view, then, we have sub-propositional thoughts which we can combine in a certain way to yield propositional ones such as the thought that there is a horse over there, or that few of the men I know own guns. We do this, he says, by joining' ideas in our minds (IV.v.2). As Leibniz pointed out, joining in my mind the idea of *man* and the idea of *wisdom* I get the thought *wise man*, which is not the thought *The man is wise*, and the latter—which really is propositional—remains unexplained. (*New Essays* p. 396)

As though anticipating this criticism, Locke writes in section 6 that he does not stand by the term 'joining' or 'putting together', and adds: 'This action of the mind, which is so familiar to every thinking and reasoning man, is easier to be conceived by reflecting on what passes in us. . . than to be explained by words.' He has, in short, no theory about how sub-propositional items are combined to yield propositional thoughts.

What about beliefs? Like most philosophers up to about a century ago, Locke does not try to analyse the concept of belief. The only general characterisation of it in the *Essay* is this:

> The entertainment the mind gives this sort of propositions is called belief, assent or opinion, which is the admitting or receiving any proposition for true, upon arguments or proofs that are found to persuade us to receive it as true, without certain knowledge that it is so. (IV.xv.3)

Someone trying to analyse the concept of belief would not help himself to 'receive as true'; in this context Locke is merely trying to distinguish belief from knowledge. I don't doubt that if he had tried to explain more generally and deeply what belief is, Locke would have given an 'entertainment plus. . . ' analysis, explaining what it is *to believe that P* by saying that it is *to have in mind the thought that P* and also. . . something further which brings it about that one actually believes that P rather than merely 'entertaining' the thought that P. But I cannot support this suspicion by pointing to texts.

In at least one place, Locke leaps over both of these hurdles, from sub-propositional to propositional, and from entertained to believed. Early in the *Essay*, at a stage where only elementary, unprocessed, un-'joined' ideas have been introduced, and have sometimes been called 'perceptions', Locke writes: 'The mind has a power in many cases to revive perceptions which it has once had, with this additional perception added to them, *that it has had them before*' (II.x.2). At this stage in his exposition he has not entitled himself to the form 'perception that P' where P is propositional.

Although a propositional thought is, in some sense, made up of sub-propositional components, it does not follow that the best way to explain what it is to have a propositional thought is through an account of some operation on sub-propositional thoughts. And although propositional thought is a genus of which belief is just one species—as Locke implies he speaks of items that 'produce in the mind such different entertainment as we call *belief*, *conjecture*, *guess*, *doubt*, *wavering*, *distrust*, *disbelief*, etc.' (IV.xvi.9)—it does not follow that the best way to explain what it is to believe that P is in terms of entertaining the thought that P and doing something further with it which marks belief off from the other species in the genus. These things that don't follow are indeed not true, according to contemporary 'functionalist'

theories of mind. These theories start with the notion of belief; and if they say anything about the genus 'entertaining', or about subpropositional thoughts, it is on the basis of and with help from their account of what it is to believe that P. If the procedure of these theories is the best one, then Locke's two failures were inevitable: He couldn't satisfactorily go from ideas to propositions, or from those to beliefs, because in each case that is the wrong order.

The thesis that propositional items are in a certain way more basic than sub-propositional ones was assumed by Kant, when he derived his list of twelve privileged concepts from a list of twelve privileged kinds of proposition. It was first explicitly declared and employed by Frege, and has had some currency ever since. The primacy of belief in the philosophy of mind became current much more recently, through the 'functionalist' view that an account of the contentful or that-P-involving aspects of the mind should start with the role that the concepts of belief and desire play in explaining behavior. It is an essential part of this position that belief and desire must be introduced and explained together: there is no chance of starting with either one and then later introducing the other. Nothing remotely like this seems to have occurred to Locke or to any of his contemporaries. Of course he knew that beliefs and desires jointly lead to action (see II.xxi); what did not occur to him, or to anyone until about a century ago, is that one might use that fact as a point of entry into an explanation of what belief and desire are.

3. Volition

Locke's treatment of desire is one theme in the longest chapter in the *Essay*, entitled 'of Power'. Its dominant theme is the issue about whether and in what sense the will is free. This is a seminal document in the literature of compatibilism: Locke argues at great length that the truth of determinism is consistent with everything that we reasonably believe about ourselves: the crucial question is whether 'the man is free' and that can be answered Yes consistently with determinism. Briefly, a person is free if there are no impediments to his doing what he wants or chooses to do, and, Locke says, there is no further problem about whether the person is free *in* his wants or choices. Many people have thought that there is such a further problem, and Locke offers several suggestions about what they might have in mind, and dispatches each of them briskly. For example, he says, they may think that the needs of morality and human dignity are not met unless *the will* is free, to which Locke replies that since the will is a *faculty* = an *ability* and not a *thing*, it makes no sense to say or to deny that it is free. (It is no accident that one of the first publications by Gilbert Ryle, who popularized the notion of a 'category mistake', was a monograph on Locke.)

Nested within this discussion are twenty pages of a different kind, in which Locke advances a theory about how, or by what, the will is determined. This is an all-purpose theory about what prompts people to act voluntarily. Of course people have all sorts of reasons for their actions, but Locke thinks that all the motivating circumstances have something non-trivial in common, and that he knows what it is: all voluntary actions proceed from some 'uneasiness' that the person is trying to relieve.

It is pretty clear that Locke thought that this was an almost obvious truth. The underlying thought is this: When I act I am trying to bring about some state of affairs S, and my trying to do that is unintelligible unless I am dissatisfied with my present non-S condition. My awareness that the non-obtaining of S is unsatisfactory to me is my uneasiness—it's my sense of something wrong—and my action is an attempt to cure it by making S obtain. For example, if I walk to the

other side of the room, that must be because I prefer being there to being here; so my present location is less than ideal from my point of view; in Locke's terminology, that means that my present location makes me 'uneasy', and so I try to relieve the uneasiness by moving.

Leibniz saw that there must be something wrong with this (*New Essays* pp. 188f). If voluntary action must always be an attempt to cure an unsatisfactoriness in one's present condition, the peak of satisfactoriness would involve perfect inactivity; but we all know that inactivity is a great source of misery. As his own rival theory shows, however, Leibniz did not get to the root of the trouble, which is this. Granted that voluntary actions must reflect a preference for some possible future over x, the relevant value of x is not *the present* but *some other possible future*. Sometimes, for example, one acts so as to bring about a future that will be just like the present in some satisfactory respect.

Locke evidently attached importance to his 'uneasiness' theory of action. Why? What did he think it does for him? Well, in the first edition of the *Essay* he advanced a different theory, namely that volitions proceed from perceptions of what is good or, rather, of what would be good if it happened.[1] By the second edition he had permanently changed his mind about this, and had come to think that a mere perception or appearance of belief about what is good cannot of itself rouse a person to volition or action. His first-edition handling of 'the greater good' made the determinant of volition and action purely cognitive, and Locke seems to have come to think that this can't be right and that something specifically conative—something motivational—must be added. This motivational item is *uneasiness*:

> To return then to the enquiry, *What is it that determines the will in regard to our actions?* And that upon second thoughts I am apt to imagine is not, as is generally supposed, the greater good in view; but some (and for the most part pressing) uneasiness a man is at present under. This is that which successively determines the will, and sets us upon those actions we perform. This uneasiness we may call, as it is, desire; which is an uneasiness of the mind for want of some absent good. All pain of the body of what sort soever, and disquiet of the mind, is uneasiness. (II.xxi.30)

Locke seems to regard his original story not as wrong but rather as incomplete: it omitted the vital link between beliefs about good and volition. Thus: 'Good and evil, present and absent, 'tis true, work on the mind. But that which immediately determines the will from time to time in every voluntary action is the uneasiness of desire fixed on some absent good.' (II.xxi.33; also 35). Note the word 'immediately'. Notice also that when Locke is arguing that his account of freedom gives us everything we can reasonably want (especially in section 48), he emphasizes thoughts about good, and not uneasiness, as a determinant of our volitions. This is evidence that he thinks of uneasiness as an addition to his previous theory, not a replacement of it.

Locke has some empirical reasons for rejecting the first-edition theory. In particular, he thinks that it is contradicted by the facts about how people will do things which they believe will prevent them from attaining infinitely great goods. (See xxi.56ff.) But he also thinks that the theory virtually stands to reason, as I have explained.

[1] The first-edition version of II.xxi.28–28 runs along the bottoms of pp. 248–273 in the Nidditch edition. The crux is 'The greater good is that alone which determines the will' (foot of p. 251), and 'The preference of the mind [is] always determined by the appearance of good, greater good' (foot of p. 256).

Where does desire fit into all this? Locke sometimes identifies it with uneasiness (II.xx.6, xxi.31,32), but that seems not to be his considered, confident opinion. He writes: 'All pain of the body. . . and disquiet of the mind is uneasiness. And with this is always joined desire, equal to the pain or uneasiness felt; and is scarce distinguishable from it. (xxi.31) The expressions 'joined' and 'scarce distinguishable' rule out an identification, although Locke goes straight on to muddy the waters by saying: 'For desire being nothing but an uneasiness in the want of an absent good. . . '.

In just one place Locke clearly implies that uneasiness causes desire: 'Wherever there is uneasiness there is desire. For we constantly desire happiness, and whatever we feel of uneasiness, so much, 'tis certain, we want [= lack] of happiness' (xxi.39). Uneasiness is unpleasant, he is implying, so one desires to be quit of it. I'm sure that this is not Locke's principal theory about how desire relates to uneasiness. If it were, he would be confronted by the question: When I want to swim half a mile and then drink capuccino and talk philosophy, how do I know that that's what I want? According to the present theory, what I most immediately want is to rid myself of a state of uneasiness, but I count as wanting those other things because I know that getting them is the way to get rid of this particular uneasiness. How do I know what the cure is? There are possible answers to this, but no plausible ones.

Locke's best and probably his most considered view is that states of uneasiness are caused by desires. That is suggested but not quite asserted here: 'Envy is an uneasiness of mind, caused by the consideration of a good we desire. . . ' (xx.13). Just what that means depends on how we take 'the consideration of a good we desire'. The following passages, however, are unambiguous: 'It raises desire, and that proportionably gives him uneasiness, which determined his will. . . ' (xxi.56). 'Good, the greater good, though apprehended and acknowledged to be so, does not determine the will until our desire, raised proportionably to it, makes us uneasy in the want of it' (xxi.35).

I believe that Locke wants to say not merely that unsatisfied desires cause uneasiness but further that that is how they cause acts of the will and thus actions. He seems to make no real distinction between desire and beliefs about what would be good; and he is saying that it/they can be effective in causing volitions only through the mediation of states of uneasiness. If there were desires (or beliefs about what would be good) that somehow failed to generate uneasiness, those desires would have no effect on action.

We should applaud Locke's seeing that he had a problem here—the problem, namely, of explaining how a mental representation of a future state of affairs can have effective power over a person's behaviour. It is typical of the depth and thoroughness of much of his thought that he doesn't rely complacently on the idea that *of course* desires contain propositions and *of course* they generate action, and instead tries to explain how these two facts are connected. He cannot be said to have succeeded, though. To do so, I believe, he would need to start again in the spirit of twentieth century functionalism, mentioned at the end of the preceding section. That would involve starting with the idea of beliefs as explainers of behaviour, and thus as collaborators with desires; there would be no notion of static belief, of something merely believed and having no bearing on conduct, except as derivative from beliefs that have a role in guiding behaviour. Although this approach was not fully developed until the past couple of decades, it was clearly adumbrated in F. P. Ramsey's suggestion that a belief is 'a map. . . by which we steer' (see D. M. Armstrong, *Belief Truth and Knowledge* (Cambridge University Press, 1973), p. 3).

It must be stressed, however, that this fruitful approach in which belief and desire are run in a single harness is hardly workable in the context of Lockean property- and concept-dualism. It is hard to put functionalism to work except as a form of materialism, namely the thesis that mentalistic facts are a subset of physicalistic facts, e.g. to have a belief is to have a complex behavioural disposition of a certain kind. Locke was nowhere near to accepting that.

4. Thinking matter

Although he follows Descartes in his dualism of properties, Locke does not confidently accept a dualism of substances. That is, he holds that there is a radical separation between properties having to do with mentality and ones having to do with materiality, but unlike Descartes he thinks that a single thing *could* have properties of both kinds. As for whether any single thing *does* have both kinds of property: Locke offers 'Do any material things think?' as a prime example of a question to which we probably cannot ever know the answer. He ignores Descartes's arguments for answering No.

In II.xxiii.15–18 and 22–32 Locke defends the notion of an *immaterial* thinking substance, but this does not seriously conflict with his latter defence of the possibility of material thinking substances. In that Book II discussion, Locke is not taking it for granted that there are thinking things and asking whether they are extended or not. Rather, he is facing up to the radical materialist—Hobbes, perhaps—who questions the entire category of *thought*, and is arguing that there are indeed thinking things. He does not and need not argue that the thinking things are immaterial. He does often say that they are immaterial, using that adjective fourteen times; but twelve of those occurrences were added in the fourth edition of the *Essay*. Michael Ayers has suggested to me that they may have been a nervous response to Bishop Stillingfleet's accusation, a year earlier, that Locke was a materialist. They muddy the waters, and should be ignored.

It is in one long section in Book IV that Locke does, taking for granted that there are thinking things, confront the question of whether or not they are extended (IV.iii.6). The notion of matter that thinks is hard to swallow, he admits, but the notion of real thing that has no extension is equally difficult to choke down, so that the reasonable stance is that of the agnostic:

> He that considers how hardly sensation is, in our thoughts, reconcilable to extended matter; or existence to any thing that hath no extension at all, will confess that he is very far from certainly knowing what his soul is. . . He who will. . . look into the dark and intricate part of each hypothesis will scarce find his reason able to determine him fixedly for or against the soul's materiality. (IV.iii.6)

We are not told what the difficulty is about real unextended things. Let us focus on the other side of the dilemma. Locke says that thought—or anyway sensation—is 'hardly reconcilable to extended matter', suggesting that there is almost a contradiction in the notion of thinking matter. But his property or concept dualism implies that there are no entailments or contradictions between mentalistic and materialistic concepts or properties, so that any description of a substance *qua* extended substance must leave logical room for the addition of mentalistic items to the description.

Sometimes, Locke virtually says as much, as in his remark that solidity and thought are 'both but simple ideas, independent from one another' (II.xxiii.32). He shouldn't have said that the ideas of thought and solidity are 'simple' in his sense: on his own showing, solidity is a 'mode', which means that it is logically complex. Still, his dualist foundation implies that they are logically non-overlapping

and thus *simple relative to one another*, as one might put it. It follows that there cannot be conceptual trouble in the idea of a thinking solid thing; and in his correspondence with Stillingfleet Locke comes close to arguing like that.

When Locke says that it is hard to 'reconcile' thought with matter, he probably means only that it is hard to see how a thing's thinking could be connected with its physical properties. Even with a severe logical separation between the mental and the physical, there still remains the question of whether an animal's material nature has some causal, less than absolutely necessitating, connection with its thought. Locke doubted that: he speaks of our 'finding not cogitation within the natural powers of matter'. But he doesn't infer that matter does not think, because he holds that it might think through divine intervention rather than through its own natural powers (IV.iii.6).

5. Dependence of mind on matter

The issue about thought and the 'natural powers' of matter is the question of whether mental facts depend on physical ones, that is, whether all mental changes are matched and causally explained by corresponding physical changes.

Locke has no Cartesian scruples about causal interaction between mind and matter. We have seen him allowing that mind acts upon matter, and he has no objection in principle to allowing causal flow the other way. But how far if at all bodily changes *do* change minds is something he prefers not to go into.[1] Early in the *Essay* he says that he won't 'meddle' with such questions as

> . . . by what motions of our spirits or alterations of our bodies we have come to have any sensation by our organs or any ideas in our understandings; and whether those ideas do in their formation, any or all of them, depend on matter or no. (I.i.2)

He seems not really to be agnostic about whether ideas of sensation depend purely on bodily states. He writes: 'Ideas in the understanding are coeval with sensation; which is such an impression or motion made in some part of the body as produces some perception in the understanding' (II.i.23). He says that we can't know whether my qualia are like yours 'because one man's mind could not pass into another man's body to perceive what appearances were produced by those organs' (II.xxxii.15). And he says that I cannot perceive an external thing except through some spatial contact with my body, because all material causation is through impact—a line of argument that presupposes that I can't perceive anything unless I am caused to do so by some change in my body (IV.ii.11).

Still, none of that implies a complete dependence of the mental on the physical; and Locke really does hold off from assenting to that. He says (and how could he deny it?) that there is probably a partial dependence in mental areas other than that of ideas of sensation:

> Whether the temper of the brain make this difference [to memory], that in some it retains the characters drawn on it like marble, in others like freestone, and in others little better than sand, I shall not here inquire, though it may seem probable that the constitution of the body does sometimes influence the memory; since we oftentimes find a disease quite strip the mind of all its ideas. (II.x.5; see also II.xxvii.27)

But this carefully stops short of complete dependence, and it is clear that Locke meant to do so. The thesis of complete dependence was a matter of anxious debate in the seventeenth

[1] The main texts are *Essay* II.x.5 and IV.x.5–6, 10, 16–17. See also II.i.15.

century. Leibniz famously denied that mental events could be causally explained in terms of events in the brain:

> Perception. . . cannot be explained on mechanical principles, i.e. by shapes and movements. If we pretend that there is a machine whose structure makes it think, sense, and have perception, then we can conceive it enlarged, but keeping to the same proportions, so that we might go inside it as into a mill. Suppose that we do: then if we inspect the interior we shall find there nothing but parts which push one another, and never anything which would explain a perception. (*Monadology* 17; see also *New Essays* pp. 66–67.)

This relies on the assumption that all physical causation is through impact, that the small differs from the large only in size, and that impact alone could not suffice to explain thought. These are tendentious assumptions. It is especially regrettable that Leibniz does not explain or defend the third.

Locke reached the same conclusion through a better argument than Leibniz's. Sometimes he treats his view about this as obvious (see IV.x.10), but in one of the places where he asserts that mentality couldn't be caused to come into existence in a non-mental world purely through a change in the material arrangements, he claims to 'have proved' this (IV.iii.6). Actually, the 'proof' occurs seven chapters later, in IV.x where Locke discusses the existence and nature of God.

Having argued that there has from all eternity been a thinking being which is the source of all other thought in the universe, Locke then considers whether that being could be material. After rejecting certain versions of that idea, he comes at last to this: 'It only remains that it is some certain system of matter duly put together that is this thinking eternal being' (IV.x.16). He means this as the thesis that the universe contains thought because, and only because, a certain material system has a structure and mode of operation that cause it to be a thinking thing. The operations of this structure must be purely mechanistic, with no help from a thinking interferer; this is because we are discussing a theory about the origin of *all* mentality in the universe: if there are any designers or guardians that must be as a result of the workings of the material system we are now discussing, and so they cannot help the system to work in the first place.

Locke argues that no system of matter could pull off this feat. His argument bears not only on whether God is a material thing, but also on what for many of us is a more interesting question, namely whether mentality could completely depend on the behavior of unaided matter. The argument is a *reductio*, starting from the hypothesis that a certain material system causes itself to have thought which is the source of all other thought. In that case, says Locke:

> If it be the motion of its parts on which its thinking depends, all the thoughts there must be unavoidably accidental and limited; since all the particles that by motion cause thought, being each of them in itself without any thought, cannot regulate its own motions, much less be regulated by the thought of the whole, since that thought is not the cause of the motion (for then it must be antecedent to it, and so without it), but the consequence of it, whereby freedom, power, choice, and all rational and wise thinking or acting will be quite taken away. So that such a thinking being will be no better nor wiser than pure blind matter; since to resolve all into the accidental unguided motions of blind matter, or into thought depending on unguided motions of blind matter, is the same thing; not to mention the narrowness of such thoughts and knowledge that must depend on the motion of such parts. (IV.x.17)

This argument, whatever it is doing, patently does not assume that seventeenth century impact mechanics must be the final truth in physics, or that the laws governing the very small must be the same as those governing the large; so it has two advantages over Leibniz's argument. But how does it work?

The argument can be seen as saying that there is some kind of regularity or orderliness such that:

(1) thought that is worthy of the name must have it,

(2) something that has it cannot be caused by something that lacks it, and

(3) no movements of bits of matter can have it unless they are under the guidance of thought.

To evaluate the argument, we have to know what kind of regularity Locke has in mind. It cannot be merely: regularity. Locke knew perfectly well that there are regular, orderly systems of matter that are not guided by minds—clocks, for example. Nor can it be: a very high degree of ordered complexity, or anything like that. Locke must have known that the ordered complexity of a material system's behaviour depends purely on the ordered complexity of its structure, and Locke seems not to believe there is any principled upper limit on that: he implies only that it is not '*probable*. . . that a blind fortuitous concourse of atoms, not guided by an understanding agent, should *frequently* constitute the bodies of any species of animals' (IV.xx.15; my emphases). The possibility of one such occurrence would be enough to kill the God argument on this interpretation of it.

If the argument is to survive, Locke must have in mind some *kind* of regularity. The only plausible candidate I can discover is the kind *teleological*. Then the argument would run as follows.

(1) Mentality essentially involves teleology: it's because the mind reaches out to possible futures that it leads people to do things so as to bring about various upshots, thus endowing them with 'freedom, power, choice'; the teleological nature of mentality is the source of the possibility of 'rational and wise thinking [and] acting'.

(2) There cannot be anything goal-oriented about the movements of matter that is not guided by thoughts, the 'accidental unguided motions of blind matter'. Therefore

(3) no such movements could be a sufficient cause for mentality.

That argument is valid, and many philosophers today would endorse its first premise: the kind of mentality that is in question here rests on belief and desire; belief alone cannot do the job; and desire is essentially teleological. But it now seems that the second premise is false: although work remains to be done on this, it is widely and rightly believed that there can goal-pursuing, teleological behavior that is mechanistically explainable. (See, for example, D. C. Dennett, *Brainstorms* (M.I.T. Press, 1978); J. Bennett, *Linguistic Behaviour* (Cambridge U.P., 1976.)

I do not claim that Locke presented his argument against dependence in full consciousness of what he was up to. When he explains that all animals have perception while no plants do, he comes close to saying that the apparent teleology of plants is not genuine, but he does not quite say it explicitly, as one would expect if he consciously held that teleology suffices for mentality. (See II.ix.11.) As for its being necessary for mentality: we have seen that Locke expends a lot of energy on a theory of volition which seems to aim at reducing the role of teleology, or at least of teleological effectiveness, in his account of the human condition.

Still, the God argument seems to have been guided by the subliminal thought that matter cannot cause teleological

patterns which are necessary for thought. If not, I do not know how the argument is supposed to work.

6. Minds and substances

Locke's famous account of personal identity (II.xxvii.9–29) is really an account of what it is for a single mind to last through time, or for two mental events to be episodes in the life of a single mind. His brilliant account of 'same plant' is extended to 'same animal', which he takes to cover also 'same man'. Or, rather, that is how he understands 'man' at the start of II.xxvii.8; at the end of the section he seems to allow 'man' to involve mental as well as animal identity, this probably being a carry-over from the account of 'same man' that he gave in the first edition of the *Essay*; and there are further complexities in sections 21f. (For details and discussion see E. Curley, 'Leibniz on Locke on Personal Identity', in M. Hooker (ed), *Leibniz: Critical and Interpretive Essays* (University of Minnesota Press, 1982).)

His treatment of sameness of 'person', on the other hand, is conducted entirely in mentalistic terms. For Locke, a man is not the same as a person. Is the man now walking past my door the man I talked to at noon yesterday? That depends on—and only on—whether there is the right kind of animal continuity linking yesterday's man and today's. But, according to Locke, whether the person now walking past my door is the person I talked to at noon yesterday depends on a mental link that has no conceptual tie to animal continuity. Even if was just one man, it might have been two persons, and it is also not absolutely impossible that it should have been different men and the same person. Because of the way it centres on mental linkage, Locke's treatment of personal identity is really an account of what it is for the mind that has thought x at T_2 to be the mind that had thought y at T_1.

Locke prefaces his treatment of personal identity with a discussion of the identity of atoms, plants, and animals. For each kind K of item, he starts with a *synchronic* account of what a K is—one that omits to say what it is for a K to last through time, that being a *diachronic* account of what a K is. In each case, he purports to infer the diachronic account from the synchronic one; the inferences are not rigorously valid, but perhaps they were not meant to be. The discussion of personal identity starts in the same way, with a synchronic statement about what a person is:

> It is a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing in different times and places; which it does only by that consciousness which is inseparable from thinking, and as it seems to me essential to it; it being impossible for anyone to perceive without perceiving that he does perceive. (II.xxvii.9)

Having emphasized the essentialness of thought to personhood and of self-consciousness to thought, Locke goes on to imply that unity of consciousness is necessary and sufficient for personal identity through time:

> Since consciousness always accompanies thinking, and 'tis that that makes everyone to be what he calls self, and thereby distinguishes himself from all other thinking things, 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. (Ibid.)

This is another attempt to get the diachronic account out of the synchronic one; it doesn't work very well, but Locke has more than that to say in defence of his diachronic account of personal identity.

The diachronic account, in effect, treats an enduring person as a special kind of aggregate of person-stages (not that Locke uses that terminology). Unlike Hume, Locke does not treat each person-stage as a special kind of aggregate of sub-personal items. Hume conceptually builds up a mind that lasts through time first by assembling mind-stages out of 'perceptions' and then assembling minds out of mind-stages. The former step is omitted by Locke, whose account of what a person starts with 'a thinking intelligent being', with no suggestion that such an item might be built up out of sub-personal items and no attempt to say what it is for two synchronous thoughts to belong to the same person.

The most powerful reworking that anyone has done of Locke's account of personal identity is more radical than Locke in just this respect. I allude to a paper in which H. P. Grice gives a broadly Lockean account of what makes two 'total temporary states' count as differently dated states of a single person, but unlike Locke tries also to say what it is for two states to be synchronous states of a single person. (H. P. Grice, 'Personal Identity', *Mind* 50 (1941).) Locke, in contrast, seems to regard the unity and singleness of *a person at a time* as a primitive, not as an upshot of how certain sub-personal components relate to one another.

He does, however, treat an enduring person as an aggregate of person stages. He devotes twenty sections to two things: a barrage of arguments against basing sameness of person on sameness of thinking substance; and the development of a positive view about what does make the different temporal stages hang together as stages of a single person.

In denying that sameness of person requires sameness of substance, Locke implies that persons are not substances. Yet his basic meaning for 'substance' is just that of 'thing', and he says firmly that a person is a thinking thing. This, as Thomas Reid pointed out, seems to be a contradiction. The trouble spreads further. The diachronic identity of an oak tree does not involve sameness of substance, Locke says, because one oak can have atoms flowing into and out of it throughout its lifetime; but it is clear that in other contexts he would classify a tree as a substance. (See for instance II.xxiii.3 and 6.)

Evidently, in this context an atom is a substance and a tree is not. It seems that here, though not elsewhere in the *Essay*, a 'substance' is a *basic* kind of thing. The general basic/non-basic distinction has several species, of which the one that is most likely to be relevant here is the simple/composite distinction: Trees are not substances for the reason that Leibniz said they are not, namely that they are composite, or have parts.[1] Trees, we might anachronistically say, are quantified over by Locke only at a non-basic level of his metaphysic, whereas material substances, atoms, belong on the ground floor. He is saying that whatever the basic, non-aggregate things that think may be, there is no strong reason to believe that a single person involves just one of these. A single enduring person might consist in or result from a steady flow through of thinking substances, as an oak tree involves a flow of atoms.

In the case of Locke's oak tree, we know what the underlying reality is that we conceptualize by saying 'this is the same oak'—i.e. we know what is involved in an oak's enduring—so we know that it doesn't involve the persistence of one or more substances. In the case of the enduring person, on the other

[1] The generic idea is explored in W. P. Alston and J. Bennett, 'Locke on People and Substances', *Philosophical Review* 97 (1988), pp. 25–46. The simple/composite species is developed in V. Chappell, 'Locke on the Ontology of Matter, Living Things, and Persons', *Philosophical Studies* 60 (1999), pp. 19–32.)

hand, we don't know what the underlying reality is. For all we know, Locke says, it may be that each person does in fact involve a single thinking substance throughout his or her existence; and he declares that 'the more probable opinion' is that personal identity involves 'one individual immaterial substance' (II.xxvii.25). I cannot find any solid basis he could have for this opinion, and perhaps Locke cannot find one either, for he goes straight on to say: 'But let men according to their divers hypotheses resolve of that as they please.' His 'more probable opinion' may have been merely an attempt to placate the indignant conservatives among his readers.

Notice: 'one individual *immaterial* substance'. Locke ought to allow—and IV.x.15 suggests that he would allow—the possibility that personal identity should be carried instead by a single material substance, an atom of matter which remained in the person's animal body amidst all the flow-through of other atoms. Locke would say that God could if he chose endow an atom with the ability to think, and could enable one atom to carry the mental history of a single person. But this seems not to be a possibility that engaged his attention.

As for the more plausible supposition that what thinks is (a part of) an animal: Locke sees that if this is right then sameness of person certainly does not involve sameness of substance:

> Those who place thought in a purely material, animal constitution, void of an immaterial substance... conceive personal identity preserved in something else than identity of substance; as animal identity is preserved in identity of life, and not of substance. (II.xxvii.12)

If it is indeed animals that think, and given that Locke is right about animals, is there any such thing as a thinking substance? Consider a thinking animal at a moment—abstracting from questions about persistence through time, i.e. diachronic identity—and ask: Are we here confronted by a momentary stage of a thinking substance? I think that Locke would say No, on the grounds that an animal at a moment doesn't constitute a substance at a moment because it is an aggregate and thus is not basic. But I am not sure about this, and can find no evidence that Locke asked himself this question.

Although he seems to hold that 'One person, one substance' is reasonably tenable only if the substance is immaterial, Locke firmly denies the converse conditional (section 12). The mere hypothesis that persons essentially involve immaterial substances doesn't imply that each person involves just one such substance, he says, unless we can 'shew why personal identity cannot be preserved in the change of immaterial substances or variety of particular immaterial substances'. He is suggesting that a person might be like a monarchy in which different kings reign, one at a time, or like a committee in which the power is exercised at each moment by a number of members. Or, of course, it might be like both at once, which would perfect the comparison with how a tree relates to its constituent atoms.

7. The same mind

So much for what personal or mental identity conceptually *isn't*. What, according to Locke, *is* it? Well, he says that the identity of a person (or a mind) through time depends upon some kind of unity of consciousness. He seems to be sure that this account best fits the plain thoughtful person's intuitions on this topic. Here, for example, we are apparently expected to find the line of thought intuitively irresistible:

> Though the same immaterial substance or soul does not alone... make the same man; yet it is plain that consciousness, as far as ever it can be extended,

> should it be to ages past, unites existences and actions [which are] very remote in time into the same person, as well as it does the existence and actions of the immediately preceding moment: so that whatever has the consciousness of present and past actions is the same person to whom they both belong. Had I the same consciousness that I saw the ark and Noah's flood, as that I saw an overflowing of the Thames last winter, or as that I write now, I could no more doubt that I who write this now, that say the Thames overflowed last winter and that viewed the flood at the general deluge, was the same self, place that self in what substance you please, than [I could doubt] that I who write this am the same myself now whilst I write this that I was yesterday. (II.xxvii.16)

This seems to rely on the thought: I have recollections of such and such experiences; what grounds do I have for regarding those experiences as mine other than that I now recollect them, i.e. the fact that there is a single consciousness that takes in both them and my present conscious state? 'If we take wholly away all consciousness of our actions and sensations, especially pleasure and pain and the concernment that accompanies it, it will be hard to know wherein to place personal identity.' (II.i.11)

In one way, Locke's analysis of personal identity is too strong, because it implies that the person who is F at T_1 is not the person who is G at T_2 unless the person who is G at T_2 does at T_2 recall having been F at T_1. (There is virtual unanimity among readers of Locke that what he calls unity of consciousness between a later time and an earlier is just episodic memory.) That makes personal identity much too tight to fit our normal ideas and intuitions about it, because we know perfectly well that people forget things that they have experienced.

As Butler and Reid saw, this feature of the analysis even interferes with the transitivity of identity: these are plenty of cases where the theory implies that x is y and y is z but x is not z. In short, identity is transitive whereas any relation such as 'remembers' or 'is a memory of' is nontransitive, and so the latter cannot be the whole analytic truth about the former.

One defence against this was deployed in Grice's famous refurbishing of Locke's theory. It weakens the analysans firstly by requiring not consciousness of being F at T_1 but just consciousness of being in some state at T_1, and then further by building transitivity into it. I find it plausible to suppose that each of these was part of Locke's intent. The resultant analysis says that

> If **(i)** the person who is G at T_2 does at T_2 recall having had some mental state H at T_1, and if
> **(ii)** H at T_1 was part of the same momentary consciousness as F, then
> **(iii)** the person who is G at T_2 is the one who was F at T_1.

Add, as part of the analysis, that identity is transitive, and the worst counterexamples disappear. To get the result that the retired general is not the person who was beaten for stealing apples as a boy, we need not merely that the general now cannot recall the incident, but that he cannot recall any previous state of himself that he was in at a time when he could recall the beating, or that he was in at a time when he could remember a still earlier state he was in at a time when he remembered the beating, or. . . and so on. It would not be madly implausible to say that if the general is as cut off as that from the beating, it wasn't he who was beaten.

A second possible defence is to say only that the person who is G at T_2 is the person who is F at T_1 if the person who is G at T_2 *can* recall being F at T_1. (Or this could be added to the weakening just discussed. That is, a single analysis

could involve transitivity, co-consciousness at a single time, *and* possibility.) That sometimes seems to be Locke's actual view, as evidenced here:

> . . . have a consciousness that cannot reach beyond this new state (II.xxvii.14)
>
> Consciousness, as far ever as it can be extended,. . . unites existences and actions. . . into the same person (16)
>
> That with which the consciousness of this present thinking thing can join itself makes the same person (17)
>
> If there be any part of its existence which I cannot upon recollection join with that present consciousness. . . (24)
>
> Supposing a man punished now for what he had done in another life, whereof he could be made to have no consciousness at all. . . (26).

The modals in these and other expressions suggest that the analysis is meant to depend not on actual consciousness but on the possibility of it. That might enable it to meet a range of counterexamples to which it would otherwise be subject.

Whether it does so, and how, depends upon what kind of modal is involved. It might be logical, conceptual. But the only basis I can find for that is the meaning of 'recall' in which it is analytic that if I recall being F at T then I was F at T. This would give a kind of truth to the statement that if I wasn't F at T then I *cannot* recall being F at T, on a par with the statement that if something doesn't have three sides then it *cannot* be a triangle; and then by contraposition we get that if I can recall being F at T then I was F at T. That reading of the analysis, however, reduces it to vicious circularity: it offers to give us leverage on 'It was I who was F at T' through 'I can recall being F at T', but the latter, we find, can be known to be true only through knowing that I was F at T.[1]

So the modality in question had better be causal: The thesis will have to be that whether the person who is F at T_1 is the one who is G at T_2 depends upon what it is *causally* possible for the person who is G at T_2 to recall at experiencing at T_1.

This notion of what a mind can do at a given time would have to be a part of any account of mentality. It's a notion that Locke demonstrably has, with respect not only to what a given mind can do at a certain moment but also to its more durable capacities and incapacities. It is conspicuous in his polemic against innatism, where he says that 'Men barely by the use of their natural faculties may attain to all the knowledge they have' (I.ii.1), that what 'the souls of men. . . bring into the world with them' are not ideas but only 'their inherent faculties' (2), and that 'there are natural tendencies imprinted on the minds of men' (iii.3). All of this, presumably, is to be understood in causal terms.

Locke is not well placed to tell us much about the causal powers of mind, especially about what the intrinsic features are of the mind by virtue of which it has these powers. This is one of those matters that he is resolutely unwilling to 'meddle' with; and it essentially involves a question which he says we cannot answer, namely whether a mind-stage is a stage of an immaterial substance, of a material substance (an atom), or of an animal.

As well as seeming to be in one way too strong, Locke's analysans for personal identity is in another way too weak. It implies that if x has experience E at T_1, and y at T_2 is conscious of having E at T_1, then x is y. On one interpretation of this, it means that if

[1] This is one of several good points made in A. Flew, 'Locke and the Problem of Personal Identity', *Philosophy* 26 (1951).

> y later is in a state which bears all the internal marks of being a memory state, and which represents an experience just like E,

then... etc. That makes the thesis much too generous about personal identity, for we can make sense of the thought of my having a memory-like state containing a representation of an experience which was previously yours, not mine. On the only other interpretation, it means that if

> y later genuinely remembers having experience E at T_1,

then... etc. That makes the thesis true, but robs it of all power to elucidate personal identity; for 'y genuinely remembers having E at T_1' entails that y had E at T_1, i.e. that the person who had E at T_1 *was y*, so that the analysans has the entire analysandum nested within it.

The best way of meeting this charge of undue weakness is to modify the analysis so that it says that if

> y's state at T_2 includes an E-type representation whose occurrence in y's mind *is an effect of* the occurrence of E in x's mind,

then... etc. That could be a first step towards a causal theory of memory which, when added to the rest of what Locke has, generates a causal theory of personal identity.[1] This causal theory has, I think, a fair chance of being true, but I cannot find the least hint of it in Locke's pages. In any case, he could have presented it only in a sketchy and abstract fashion, because he declines to have any views about what kind of item a mind is.

8. The mind's continuity

Descartes held that thinking is the whole essence of minds, and extension the whole essence of matter. This committed him to two biconditionals, namely: Necessarily, for all values of x,

> x is a mind when and only when x thinks,

and

> x is a portion of matter when and only when x is spatially extended.

Locke accepts one half of each biconditional and rejects the other. Agreeing that all matter must be extended, he says that there can be extended items that are not material, namely stretches of empty space; agreeing that whatever thinks is a mind, he denies that whatever is a mind at time T must be thinking at T, i.e. that 'actual thinking is as inseparable from the soul as actual extension is from the body' (II.i.9). Even if thinking is 'the proper action of the soul', it does not follow, and is not true, that the soul is 'always thinking, always in action' (10).

That is near the start of II.i.10–19, which is entirely devoted to arguing that 'the soul thinks not always'. On this matter, Locke is content to take his stand on his own knowledge—as he thinks it to be—that last night he slept dreamlessly; during that time, he says, his soul was not thinking.

Here again we run into the question that Locke cannot answer: What kind of item is a soul or mind? When it is quiescent, or not 'in action', in what does its reality consist? As well as not answering this, I suspect that Locke did not even ask it. That is, he seems not to work with any robust

[1] For a contemporary causal theories of memory and personal identity see, respectively, C. B. Martin and M. Deutscher, 'Remembering', *Philosophical Review* 75 (1966); and J. Perry, 'The Importance of Being Identical', in A. E. Rorty (ed), *The Identities of Persons* (University of California Press, 1976). For a deeper exploration of some of the issues I have discussed here, see J. Perry, 'Personal Identity, Memory, and the Problem of Circularity', in J. Perry (ed), *Personal Identity* (University of California Press, 1975).

idea of a soul or mind or person as a continuously existing item. In his treatment of personal identity he uses such turns of phrase as 'whether the same self be continued in the same or divers substances' (II.xxvii.9), and 'continued in a succession of several substances' (10); see also 25 and 29. But I cannot find in this chapter, or anywhere else in the *Essay*, any working notion of mental continuity that goes beyond the mere possibility of reidentification of a single mind or soul or person at different times. The concept of a person could be such as to permit such a reidentification across an ontological gap; and, while I have no evidence that Locke believed that there are such gaps, nothing in his thought seems to reflect a solid conviction that there are not.

When the diachronic identity of others kinds of things is in question, it's a different story:

> •. . . an atom, i.e. a continued body under one immutable superficies. . . [It] must continue as long as its existence is continued. . . (II.xvii.3)
>
> •. . . such an organization of those parts as is fit to receive and distribute nourishment, so as to continue and frame the wood, bark and leaves etc. of an oak. . . It continues to be the same plant as long as it partakes of the same life. . . ...parts of the same plant during all the time that they exist united in that continued organization.... (4)
>
> •. . . what makes an animal and continues it the same. If we would suppose this machine one continued body, all whose organized parts were repairs [etc.] by a constant addition or separation of parts, with one common life, we should have something very much like an animal. . . (5)

For atoms, plants and animals, continuity through time is insisted upon. This is in contrast with Locke's treatment of the diachronic identity of minds, in which continuity is not mentioned and, from Locke's examples, seems not to be required. Thus, if we ask Locke to tell us how things stand with a mind when it is not thinking, e.g. when its owner is dreamlessly sleeping, it would be harmonious with the over-all tone of his philosophy of mind for him to say: 'While the man is sleeping and not dreaming, there isn't any such object as his mind or soul. The fundamental reality at that time consists in a sleeping animal which can, and when it receives certain stimuli will, start thinking again.' This is a long way short of the kind of materialism that finds favour with most Anglophone philosophers today, but it is a step along the way.

It is, furthermore, a step that can be taken consistently with the dualism of properties and concepts that Locke inherited from Descartes. Even while maintaining that form of dualism, Locke could have taken the position that there is no such item as a mind, and that colloquial uses of 'mind' are just ways of talking about the mental lives of animals.