23
Locke, *Essay* II.xxvii.1-14
Identity

Locke's chapter on the idea of identity carried implications that were even more disturbing for many of his contemporaries than were his views on substance. Over the course of the chapter, he developed a distinction between the body (or "Man" as he called it), the soul (or spirit), and the self (or person), and he argued that the self is radically distinct from both the body and the soul — so radically distinct that, in principle, the self could be conceived to change the soul in which it inheres as easily as bodies change their clothes, and souls could be inhabited by a succession of different persons over time. As Locke understood it, the self is the collection of our sensations, thoughts, desires, and volitions, whereas the soul and the body are substances. On this view the self is fragile. Rather than be intrinsically capable of existing on its own, as a substance, it is merely a complex mode. This account rendered worries about the immortality of the soul and its distinction from the body entirely moot. Even if the soul were immortal and distinct from the body, it would not follow that the self survives death.

QUESTIONS ON THE READING

1. What is requisite if we are to be sure that one thing is different from another, e.g., that someone has a twin?
2. What are the three kinds of substance?
3. Why is identity not a problem where God is concerned?
4. Are there any exceptions to the rule that two different things cannot be in the same place at the same time?
5. Under what condition would our ideas of identity and diversity have no foundation and be of no use?
6. What determines the identity of particular instances of modes or relations and distinguishes different instances of the same mode (e.g., different right-angle isoceles triangles of the same colour and size) from one another? What sorts of modes and relations can have no identity?
7. If a mass of atoms has its parts rearranged, is it still the same mass? If it loses or gains a part, is it still the same?
8. What makes an oak tree different from a mere mass of matter?
9. What distinguishes one plant from another of the same species, e.g., one oak tree from another?
10. What determines that the organization of life-function performing parts that is present at any one instant in any one collection of matter is identical to one earlier or later organization of life-functioning parts characteristic of that species of plant rather than some other, e.g., that the bush in this pot in winter is identical to the one that was in the garden in summer?
11. How does the identity of animals and plants differ from that of machines?
12. What is wrong with supposing that what makes people the same from one moment to the next is that they look the same?
13. What is wrong with saying that what makes someone the same human being from one moment to the next is that they continue to possess the same soul?
14. How do we distinguish our self from the selves of other thinking beings?

15. What determines how far my self exists backwards in time, that is, what determines whether I will consider any past thought or action to be identical with a thought or action of my self?
16. Does identity of self presuppose identity of thinking substance?
17. What grounds do we have for affirming that the same person cannot be successively present in different thinking substances?
18. What grounds do we have for affirming that different persons could be successively present in the same thinking substance?

NOTES ON THE READING

Unlike the idea of substance, which is a complex idea, our idea of identity is the idea of a relation that we find between ideas when we compare them with one another.

At its most basic, this relation is the relation of being exactly alike in all respects. We can call this strict identity, and we can say that the idea of this relation arises in us when we compare one thing with itself, and discover absolutely no differences whatsoever. For example, I might compare a chair with itself, and assert that it is identical with itself because the comparison discovers no differences.

However, Locke was worried about a different sense of identity: identity over time.

The difference between these two senses of identity can be illustrated by an example. When the victim of a crime points to one of the people in a police line-up and says: "That is the one," we say that the victim has identified the perpetrator. But the individual who has been identified cannot possibly be exactly identical in all respects to the individual who committed the crime, if for no other reason than because the individual in the line-up is older than the person who performed the crime, if only by a few hours.

We think that objects can continue to exist over time, and that an object existing at a later time can be, as we say, "the same as" or "identical to" an object existing at an earlier time. But this notion of identity over time is importantly distinct from strict identity. We do not think that a later thing has to be exactly identical in all respects to an earlier thing. Later things can look very different from earlier things and still be thought to be identical to them. Think of earlier and later states of the same fire, the same acorn/sapling/oak tree, the same egg/tadpole/frog, or the same infant/child/adult. Moreover, later things that look to be in all respects identical to certain earlier things may in fact be different from them, for example, the chairs in one classroom today, and those in that classroom tomorrow. But if resemblance does not determine identity over time, what does?

Locke approached this question by first observing that our notion of strict identity entails that no two things of the same kind can be in the same place at the same time. (He added the qualification "of the same kind" because we do think that things of different kinds can be in the same place at the same time. For instance, a colour and a taste can be in the same lump of sugar at the same time, though two different colours or two different tastes cannot. And if spirits are distinct from bodies, and spirits can be located in space, as Locke was at some pains to argue in *Essay* II.xxiii, then a spirit and a body might be in the same place at the same time.) For, when you take your idea of what exists in a given place at a time and compare it with your idea of what exists in that place at that time you see no differences whatsoever. So your idea is one, strictly identical thing, not two different things. Presumably, since it is just one thing, it can only be in one place at any later time. (Were it capable of splitting in two, it would have had to consist of two parts set outside one another at different locations in space to begin with and we should have been considering just one of those atomic parts. The case of plant and animal generation will be

considered below. For now, let us just focus on particles.) So whatever later things can be traced back to this thing must be identical with it. As he put it, whatever later things have the same "beginning of existence" (whatever later things originated from the same place at the same time as some other thing) must be identical to that thing and to one another.

Locke's proposal for dealing with identity over time is not that different from that employed by the police when attempting to identify the perpetrator or a crime. If a suspect turns out to have been in a different place at the time the crime occurred, that person is eliminated from suspicion, because no one can be in two different places at the same time. But if the investigation is able to trace the suspect's history of motion and rest right back into the footprints of the perpetrator at the time of the crime we think that the investigation has identified that suspect as the perpetrator, and that no further investigation of the other suspects is called for. For, just as nothing can be in two different places at the same time, so no two things of the same kind (in this case, no two different people) could be in the same place at the same time. So, identifying just one suspect as the perpetrator automatically eliminates all the other suspects from suspicion, even in the absence of independent exoneration.

The two principles just mentioned,

– nothing can be in two different places at the same time
– two different things of the same kind cannot be in the same place at the same time

imply that each distinct thing that now exists can be traced back to one and exactly one place, not occupied by any other thing of the same kind at each past moment of time. For, were it possible to trace two distinct things of the same kind back to the same space at some past time strict identity would be violated, since what exists in a place at a time is exactly the same as itself in all respects and so one and not two. And were it possible to trace one thing back to two different spaces at the same past time, then there would be one thing in two different places at the same time, which is impossible. From these reflections we can obtain both a principle of individuation and a principle of identity:

> *Principle of individuation*: whatever is in two different places at the same time is two different things
> *Principle of identity*: anything with a history that puts it in the same place at the same time as some earlier thing of the same kind is the same with it.

Locke laid this account out succinctly in II.xxvii.1

> That therefore that had one Beginning is the same Thing, and that which had a different Beginning in Time and Place from that, is not the same, but divers. [*Essa*y II.xxvii.1]

As thus presented, Locke's account of identity is dissatisfying. The account appeals to the notion that the way to determine whether one object is the same as another is to find out whether it had the same beginning of existence. But how could we do that? Above, I alluded to studying the object's history of motion and rest. But motion and rest are things that happen over time. To determine whether a thing is moving or at rest, I have to be able to identify where it is from one moment to the next. If I have to be able to identify things from one moment to the next in order to determine their history and motion and rest, then I can't take their history of motion and rest to be what determines their identity over time. That would get me caught in a vicious circle.

This is not an issue that Locke addressed. Perhaps he would have said that the circle could be broken by stipulating that you keep your eye or hand constantly on the object. If you wrap your hand around an object, so it gets carried along with the object wherever the object goes, then you can be sure how the object is moving over time. Derivatively, if you can think that if you had wrapped your hand around the object (even supposing that is in fact impossible), then your hand would have been carried to a certain place, then you can be assured the object in your hand is identical to the one you grasped earlier. The trouble with this approach, however, is that it is just as question begging. The question is what makes things identical over time. I am a thing too. So I need to know what makes me identical over time before I can invoke my earlier and later visual and tactile sensations in order to determine the identity of other things.

Locke might have objected that his purpose was not to answer the epistemological question of how we come to know whether or not one thing is the same as another, but only to answer the metaphysical question of what as a matter of fact makes one thing be the same as another, whether we can know it or not. But this metaphysical question is still at issue as well. It remains unresolved what makes an object considered to exist at a given place at a given time the same as some other object, existing at that place or some other place at an earlier time. Perhaps Locke would have answered this question by appeal to the laws of physics, which would allow us to take an object existing at one time to exist with an inertial tendency to move in a certain direction and at a certain speed. That current inertial tendency determines where the object will be at the next instant, and (because the laws of physics are time-reversible) where it must have been at the prior instant, correcting for collisions or other influences (such as gravitation by surrounding bodies). On this account, then, it is the momentary inertia together with currently occurring interactions with surrounding bodies that determines the body's history of motion and rest.

This is still only a partially satisfying answer, precisely because it only answers the metaphysical question of what makes things identical over time, leaving the epistemological question of how we come to think of them as being identical over time unresolved. Locke could not afford to leave this question unresolved because he considered identity to be a primarily "forensic" notion. We make use of it to determine who, among the people existing now, is responsible for crimes committed at an earlier time. This means that identity relations have to be ones we can come to know. And we can't come to know them by applying the metaphysical criterion of taking inertial states to determine a body's history of motion and rest. This is because the only way we can tell what inertial state bodies have at any given moment is by determining how they were moving at the past moments, which requires identifying them over time.

Locke's appeal to the principles that nothing can be in two different places at the same time and that two different things of the same kind cannot be in the same place at the same time at least entails that identity must be a one-to-one relation. For any given thing, there can be at most one thing at any given earlier time that is identical to it, and at most one thing at any given later time that is identical to it. Perhaps we base our judgments of identity on this constraint, together with the further principle that motion and rest must be continuous (that there can be no gaps or jumps, so a thing cannot cease to exist at one place and then pop back into existence somewhere at some later time, or leap from one place to a separate one next without first taking a path across the interval), and a reintroduced notion of resemblance. Considering very small intervals of time and space, we ask which of the objects in the small spaces surrounding the spot a given object occupied at the fractionally earlier time, is the one that most resembles the given object. In cases where there are

two or more such objects, we look yet smaller spaces at yet shorter times. In cases where this does not help, we remain genuinely confused about which object is which, but this only happens rarely.

In practice, determining identity in the way I have just proposed is tantamount to determining identity by keeping an eye or a hand constantly on the object. If it moves, at the next small moment of time you see that the thing most resembling it is at one of the adjacent small regions of space. If it stays at rest, at the next small moment of time, you see that the thing most resembling it is in the same small region of space. This test only fails when the surrounding objects are all so similar to the target object as to be indistinguishable from it but for their locations. In these cases, hopefully the things are all welded together as parts of a single thing and not in motion relative to one another — or if they are in motion (like water drops in a river), we don't care about identifying them from one moment to the next but are only concerned with the larger mass. Otherwise, we just can't determine identity over time. This shouldn't trouble us, however, because sometimes identity attributions are very difficult to make. And, as mentioned earlier, the genuinely perplexing cases are fortunately rare.

Earlier, I said that determining identity by keeping an eye or hand constantly on the object is question begging, because it presumes that we can identify ourselves over time. But there might not be the same problems with ascribing identity to myself as there are with ascribing identity to other things. If anything is obviously identical over time, we are obviously identical for ourselves over time. Of all the people in existence, there is only one whose past sensations I remember. Moreover, I am pretty confident that no one else can remember those sensations. This fact of my "privileged access" to my past sensations makes me certain of my identity with the person whose sensations I remember. So if I put my hand on an object and attend to where my hand is as the object carries it along, I can be sure that I am sensing my hand throughout, not having someone else's sensations of their hand. And, given that my hand has remained in contact with the same object (i.e., has been giving me the same sensations), I can be assured I am identifying the same object at the later time that I touched earlier. Once we have made these tactile judgments of identity with things like houses and trees, and determined that these objects do not move relative to one another over time, we can use them as landmarks relative to which we can identify unchanging locations in space, and make visual determinations of identity based on where things are located relatively to these landmarks, and where most resembling things are in the small, immediately surrounding neighbourhood at short intervals of time.

In laying out this position on how we might determine a thing's history of motion and rest I have gone beyond the letter but not the spirit of Locke's account. But we are not entirely out of the woods because Locke proceeded to draw a number of surprising conclusions, some of which challenge our confident assumption that we can identify our hands or bodies over time. Fortunately, it turns out that the possibility of such misidentifications is very remote — so remote as to be a mere logical possibility rather than anything that is physically likely. That bare logical possibility, however, poses serious challenges for other things than our ability to identify things over time.

Identity of Simple Things. Among the things that exist there might be some that are simple, meaning that they cannot be divided into parts. Unbreakable atoms would be such things. Souls, supposing that souls cannot be divided into parts, would be another.

If atoms and souls exist, then the rule for their identity over time would be a direct application of the historical criterion: if a study of the past history of motion and rest of a given atom or soul

shows it to have occupied the same place at the same time as some earlier atom or soul, then those two atoms or souls are identical, and the first is a later state of the second; otherwise, they are distinct.

However, the fact that atoms and souls are different kinds of things means that, were we to trace the history of an atom back to an earlier time when it occupied the same place as a particular soul, we would not consider the atom to be the same as the soul. Things that occupy the same place at the same time are identical only if they are of the same kind, that is, if they are both atoms or both souls.

The fact that an atom could occupy the same place as a soul without the two of them being considered to be identical means that, at a later time, the atom and the soul could move off in different directions and the atom come to occupy the same place as some other soul while the soul comes to occupy the same place as some other atom. Otherwise put, reincarnation, the act of a soul leaving its "body" and coming to occupy a series of different bodies in succession is at least a logical possibility.

Identity of Compound Bodies. The simplest way to deal with the identity of compound bodies is to take it to be a function of the identity of their component parts: if all of the parts of a compound body are identical to all of the parts of some earlier compound body, then those two compound bodies are identical; otherwise, they are distinct.

Locke did not say so, but, in practice, this criterion often proves to be too strong. If a raindrop erodes a grain of sand from a mountain, and so removes some of its atoms, we do not say that the mountain becomes a different mountain, and is no longer identical to the mountain that existed before the scrape. Of course, there comes a point where the removal of material becomes so extensive that we do not think that the same compound body still exists.

This goes to show that there is a degree of vagueness to our concept of the identity of compound bodies. There is no precise point at which identity ceases and diversity begins. But there are points at which it is obviously present (when no atoms have been removed) and when it is obviously absent (when a significant proportion of them have been removed).

However, we are not always satisfied even if exactly the same atoms continue to be present in a compound. If I have a watch and someone smashes it with a hammer into a pile of rubble, all the atoms that were originally present may continue to be there. Only their manner of arrangement has changed. Yet I am not inclined to say that the watch still exists. In this case, it is not just the continued presence of the same collection of atoms, however they may be mixed or arranged, that establishes identity. The manner of arrangement of the atoms must also persist. This is not always the case. A pile of laundry is the same pile even if the clothes making it up are jumbled about. But in many cases rearrangement of the parts of a compound is taken to destroy its identity.

In some cases, the manner of arrangement of the parts is so important that we are even willing to ignore significant changes in the atoms that go to make a compound up, as long as the manner of arrangement is preserved. This is the case with the ship of Theseus, which over the years was wrecked and repaired so many times that not one nail that was originally present in the ship remained. Yet Theseus and his companions continued to give the ship the same name and to consider it to be the same ship they had sailed out in.

As Locke understood it, these sorts of cases show that we are not always concerned with the identity of *substance* (the atoms making up a compound) over time. In many cases (and machines like ships are a good example), we are more concerned that the *manner of arrangement* of the

substances persist over time, even if the individual substantial parts arranged in this manner get exchanged with other, similarly functioning parts. (This manner of arrangement is what enables a machine to perform its function, and that explains why we might be more concerned with it than with the substances that make it up.)

The manner of arrangement of substances is a *mode*. Recall that modes, like substances, are collections of simple ideas. But whereas substances are collections of simple ideas that we come across in our perceptual experience, modes are collections of simple ideas that we only ever think of in conjunction with yet other simple ideas. A sugar cube is a substance, being cubical is a mode. The collection of ideas that goes to make up what we call a cube can only be perceived when we perceive some substance that exhibits this mode. But we may perceive different substances to exhibit the same mode.

Locke observed that modes can persist even if the substance changes. For example, if a compound body consists of a number of parts of a certain shape, arranged in a certain fashion, and some of these parts are removed and replaced by identically shaped and arranged parts, then the mode of arrangement remains the same, even though the substance has changed, for the mode just consists in the placement of shapes in a certain order, and does not include any reference to what bears those shapes. This suggests that modes are amenable to being identified and individuated in the same way as substances: We can say that any particular manner of arrangement of substances that now exists at a certain place is distinct from the identical manner of arrangement of substances now existing in some other place. And we can conceive of these different manners of arrangements as persisting over time and having histories of motion and rest that establish their identity with their own earlier and later states and their distinction from other complex modes of the same type.

This is what Theseus and his companions did with their ship. They considered it to be the same ship because what they were concerned with was the mode of arrangement of the substances, rather than the substances themselves, and that mode of arrangement had a history of motion and rest that individuated it from other ships (or ship-shapes), and that gave it an identity over time irrespective of the exchange of its parts. Had the ship been totally dismantled and then rebuilt from scratch some time later, then even if the rebuilt ship was an exact replica, made with exactly the same parts, there would have been a break in the continuous existence of that particular mode of arrangement of substances, and Theseus and his companions would not have given it the same name or considered it to be the same ship.

Identity of Living Creatures. What Locke said about the identity of machines applies to the identity of living creatures: plants, animals and human beings. But some serious modification is required. Unlike machines, which merely exchange their component substances over time as they break and get repaired, plants, animals and human beings grow, starve, reproduce, occasionally suffer accidents that result in a loss of limbs or organs, and even undergo radical changes of form (say from egg to larvae to adult insect, or tadpole to frog). Accounting for their identity therefore poses further challenges.

Locke thought that in these cases our ascriptions of identity are still based on the continued persistence of a mode over time, but that the mode is not a mode of arrangement of parts, but something rather more complex: an arrangement of parts that work together to preserve the same life. Here, “life” is something that we take to consist in the ability to perform certain functions such as respiration, nutrition, growth and reproduction, locomotion, sensation, or understanding. At any given moment a particular organization of parts that enables a compound substance to be

considered to be alive is distinguished from any other compound substance, however similar, that exists in a different collection of parts in some other place at that time. Then, whatever substances may flow into or out of that compound over time, and however the organization of its parts may change over time, as long as the same life functions continue to be performed in that compound, the identity of the plant, animal or human being is preserved. Thus, when a woman becomes pregnant, we say a second life begins in a cavity inside her body at the time when the fetus or embryo begins to perform its own life functions — functions that make it alive insofar as they are performed but that make no contribution to the life of the mother and so would not take away her life were they removed from her. Something similar can be said about the relation between all of us and parasites or disease organisms that we might contain. For the same reasons, if an amoeba splits in two, what we say is that the parent life is destroyed in giving birth to two children.

Personal Identity. Locke's reflections on the identity of modes led him to claim that though the idea of identity is fundamentally the idea of the relation of originating from the same point in space at a particular time, in different cases there are different things that we attempt to trace back to a common point of origin. When considering the identity of physical atoms, souls, and masses of matter, it is the substance that we attempt to trace back to a common origin. But where the identity of plants, animals, human bodies, machines, and other artifacts is concerned, it is not the substance that we are concerned to trace back, but modes of varying complexity — modes that may, for all we care, have been transmitted from substance to substance over the time of their existence.

Locke went on to observe that there is a further case where this happens, the case of the identification of persons. That Locke should have said this might seem strange. After all, he had already given an account of how we identify souls or minds from one moment to the next, as well as an account of how we identify human bodies from one moment to the next. What, then, is left over for the term, "person," to refer to, if not just one or the other of these two things?

The best way to grasp what Locke was worried about is to recall the position he took on the identity of human bodies. He there said that I can be considered to continue to have the same body even though all the atoms that go to constitute that body might change over time. I — what I consider to be my self — could not therefore be identified with any particular physical atom or collection of physical atoms that might at any time happen to be present in my body or to constitute it or one of its parts. Were I to identify my self with my body, therefore, I could not identify it with any atom or collection of atoms. I would have to identify it with a complex mode. Should we say, then, that I just am the complex mode that constitutes my life? That as long as the same life continues to exist I must continue to exist, regardless of how my body might change over time?

There is a problem with saying these things. Recall that earlier Locke allowed that, in principle if not in fact, souls might be conceived to be coincident with different atoms at different times. Consequently my soul could conceivably exist in different bodies. If I take my self to be wherever my soul is, then my self could be in different bodies at different times.

But even if we deny that there are such things as souls, there are reasons to suppose that the self might not always exist where the body is. A series of thought experiments can help to illustrate this point. We typically identify persons on the basis of their human bodies. For instance, when someone is hauled into court accused of a crime, we determine whether they are guilty or not principally by determining whether they have the same body as the criminal, often by appeal to characteristic traces that only a particular human body can leave behind, like fingerprints or DNA samples. But we do not really think that the body performed the crime, any more than we think

that the gun or knife performed the crime. The body is just an instrument used by the perpetrator, just like the gun or knife. Suppose that the technique of transplanting brains were perfected and that a criminal, just prior to performing some crime, kidnapped an innocent person and had his brain transplanted into that person's body, used the person's body to perform the crime, being careful to leave DNA samples and fingerprints behind, and then had his brain transplanted back into his old body. In this case we would say that even though the DNA and fingerprint evidence proves that the kidnapped person's body was the body that performed the crime, the *person* who had been kidnapped was not the same *person* who had performed the crime. Though we look in the first instance for the same body when attempting to identify perpetrators, we do this only because brain transplants are not yet feasible, let alone common. But were brain transplants common, we would not think that persons should be identified just by identifying their bodies. We would think that, over time, it would be possible for the same body to be occupied by different persons, and for the same person to move from one body to another.

Some might object to this, and say that were something like a brain transplant to be possible, then it would change the identity of the living thing, so that a body that looses one brain and gains another should no longer be considered to be the same life. But this would be a hard line to sustain. We do not think that liver transplants change the identity of the living thing, and livers are more or less the same size as brains and perform a number of functions that are necessary for life. By parity of example, brain transplants should not destroy the identity of the living thing either. Moreover, there is a sense in which brain transplants do occur and occur all the time in all of us. Through the processes of nutrition and excretion, our brains are constantly taking on new material and excreting old material. In effect, they are being transplanted, part by part, over time. Yet we do not think that this changes their identity. According to Locke's criterion for animal body identity, it does not matter what organs are transplanted into our out of the body, or what atoms are ingested or excreted. As long as the body continues to live through the operation, it is that same body. So if my brain is transplanted into your body and your brain into mine, your body is still the same body after the operation that it was before, as is mine. I might think that my self is now in your body, but then I must think of my self as something distinct and separable from my body (as I do insofar as I think that you now have "my" body).

This raises a question: Why are we inclined to suppose that our selves have changed bodies when we have a brain transplant? If a brain transplant is no different from a liver transplant, and I think that I would be the same self after a liver transplant, why do I think that a brain transplant would move my self into a different body, rather than think that I would still continue to be the same self in the old body?

The answer to this question seems to rest with the fact that, while brains and livers both perform important life functions, only brains perform cognitive functions. It is through our brains that we sense, think, feel, desire, and will. In a word, it is through our brains that we are conscious. And, as Locke pointed out, we take our self to be wherever our consciousness is. If I suddenly start seeing the world as if I were peering out through your eyes and feeling it by feeling what is happening to your body, then I think that my self is in your body.

This is also why, in the case mentioned earlier of the criminal who undergoes a brain transplant in order to perform a crime, our intuition is to charge the body that contains the criminal's brain with the crime, rather than the body that actually performed the crime. For we think that it was through the brain that the *mens rea* or criminal intention and decision to perform the criminal deed was formulated and executed.

But we should not slide into thinking that we could simply equate the brain with the person. After all, as noted earlier, the brain itself is constantly being altered over time. The atoms that go to make it up are constantly being excreted and replaced with new atoms. If it is the brain that is responsible for our consciousness, then it is not the substances that go to make it up that make us consider ourselves the same selves, but rather the way those substances are organized to enable the brain to think the thoughts it does.

But manner of organization, being a mode, is something that can be transferred from one substance to another. We could imagine an operation that does not actually transfer any substance from one body to another but merely alters the manner in which the neurons in the brain are connected, erasing the connections from one person's brain and replacing them with the connections in another person's brain. If consciousness is nothing other than just what is produced by the manner of arrangement or connection of the substances in the brain, then transferring this manner or arrangement or connection should constitute transferring consciousness.

We seem compelled, therefore, to recognize that selves are distinct from living bodies and even from brains. But the same set of considerations that we have just raised concerning the relation between selves and bodies also applies to the relation between selves and minds, supposing that such things exist. We think of minds as simple spiritual substances. But are we sure? Might minds be compounds of many spiritual substances the way bodies are compounds of physical atoms? If so, then perhaps selves or persons might be complex modes arising from the organization or functioning of compound spiritual substances. But if that were the case, then it might be possible for consciousness to be transferred from one mind to another for just the same reasons that we earlier speculated that it might be possible for consciousness to be transferred from one brain to another.

Though we do not know what makes up our selves — whether they are simple spiritual substances, compound spiritual substances, or somehow constituted from physical substances — we do know the functions that these selves perform: each of us is aware of his or her own experiences and past experiences, and whatever the self may be, wherever that awareness is, that is where the self is.

This awareness or consciousness can be used as a criterion for personal identity. For one of its features is that it just is "personal." Only I can experience my thoughts and remember my past experiences. No one else can know what is going on in my consciousness, unless of course I choose to tell them, or unless I engage in certain behaviours that allow other to conjecture what I am thinking, and even then what others know is not my thoughts, but really just my words or actions, from which they infer my thoughts. Similarly, no one else can remember my thoughts. The best that they can do is remember that I said or did things, and infer from that that I likely had certain thoughts.

This is why Locke declared that the identity of persons is determined by the extent of their consciousness. Those thoughts, sensations, desires, volitions, and memories that I am now conscious of constitute my current person. Any other, currently existing thoughts, sensations, desires, volitions and memories that I am not conscious of must constitute other persons. And of all the past persons that existed at any given past moment in time, there is at most one whose consciousness I may have some partial access to, via memory whereas no one else has the least access to that person's consciousness. This makes me identical with that past person.

But my thoughts, intentions, volitions, and memories — the things that go to make up my consciousness — are really just bits of information. Today, we think that what makes a bit of

information the same from one moment to the next is just that it continues to say the same thing. The identity of a bit of information has nothing to do with how it is stored or retrieved. A bit of information may be stored in a book, on paper; in words, on magnetic tape; or in "bits" and "bytes" on a computer disk. It is the same bit of information regardless of where it is stored (or what substance it is stored on).

In contrast to consciousness and memory, which are bits of information, brains and souls are substances or, as we would put it today, storage systems for bits of information. And in principle, just as information can be deleted from one storage system, e.g., one computer hard drive, and transferred to another, so consciousness and memory, being just a string of bits of information, can be transferred to a different storage system — a different brain or soul — and still be the same. They could even be copied to a number of different thinking substances simultaneously and wiped clean from the original thinking substance that they were imprinted on. Or a thinking substance that has had its consciousness and memories wiped away could subsequently have them recopied, with new information added.

Locke did not know about the distinction between hardware and software in computers or have a concept of information that can be copied from one storage system into another. But his concepts of thought and will as modes and minds or brains as substances did the same job for him. They led him to hold that if we are worried about identifying the person who performed a crime, that is, the consciousness that had the criminal act of will, then we need to worry about where the modes (thoughts and volitions) are, not where the substances (minds or brains) are.

By this criterion, the person who performed the crime is the one who is now consciousness of having formed the criminal intent and willed the criminal act. And the person who now remembers having performed that criminal act of will is the guilty person, regardless of which soul or which body that person may now have come to inhabit. But then, by the same token, if that person should have been wiped out or destroyed, somehow, then even if the body and indeed even if the soul that the person inhabited should persist, it could not justly be held responsible for the crime. If a criminal convicted and sentenced to be executed is shot and killed while attempting to escape custody, and the criminal's heart is transplanted into someone else, the justice system does not demand that, on the date set for the execution, the heart be removed from this other person for execution. Similarly, if the criminal consciousness is no longer in the whole body or the whole soul it would be equally unjust to proceed with the execution. Consider a case where a person convicted of a crime suffers a blow to the head and loses all memories of his or her past life. In this case we can be sure that the body has persisted, but how can we tell whether the self or person was not destroyed by the accident? The only thing we have to go on as a criterion in determining the identity of selves is their consciousness of their own past states and where that is lacking we are in no position to claim that the self persists. This is why Locke claimed that it would be wrong to punish sleeping Socrates for crimes that waking Socrates performed because sleeping Socrates has no more knowledge of that than you or I do.

Of course, we have no way to see into the souls of others to determine whether they remember performing criminal acts. This means that when attempting to identify perpetrators of crimes we have to fall back on other, less reliable criteria, such as having the same body or the same brain. We do this in the belief (usually true), that souls do not get translated from one body to another, and consciousness does not ever get fully erased. But the importance we attach to the idea that it is necessary to identify the consciousness that performed the criminal act of will is revealed by our wanting to obtain confessions from criminals, even when the evidence against them is otherwise

clear, and by our reluctance to punish criminals if, for whatever reason, they have no memory of the criminal act. Think, for example, of cases where individuals who performed crimes while sleep-walking have been acquitted of guilt, even though their bodies clearly performed the offense.

Implications of Locke's position on personal identity. Locke's conclusion is that what is ultimately most important to us is not identity of substance (that is of body or soul) but identity of person. This is because it is the person who contains the consciousness of having willed, and so the person is the moral agent responsible for crimes and good deeds. It is this moral agency that we are most concerned with in our dealings with others (though identity of body is relied upon as an indication of where the identity of person is likely, but not necessarily to be found). Identity of soul-substance is actually irrelevant for determining the identity of moral or responsible agents. For, firstly, soul-substance cannot be perceived, whereas the identity of our own consciousness can be perceived by ourselves, and that of others reported to us by them. And, secondly, in principle, consciousnesses could be transferred between different souls, and since it is the consciousness that performed and remembers the criminal or noble acts of will, that is what we are ultimately concerned with.

Indeed, even for God, who knows all and is able to see where the same souls are from one moment to the next, same soul and same person are different notions. God could, if he wanted, transfer consciousnesses between souls and so make one soul contain the memory of criminal deeds that were willed by another soul.

Locke drove a wedge between our concepts of personality and moral agency on the one hand, and soul or thinking substance on the other. Many people in Locke's day were deeply committed to the notion that we have an immortal soul. But if Locke is right, the thing that is morally responsible for good and evil deeds, and that is therefore deserving of reward and punishment in this life or the next, is not the soul but the person. And far from being immortal, the person is constituted of one of the most ephemeral and transitory things there is: a mere modification of a substance that can easily be wiped out even while the substance remains the same. In Locke's day, this was perceived as a pernicious and dangerous teaching, that called the afterlife into question and attacked the basis for morality by leading people to doubt whether they would be punished in hell for their misdeeds.

Locke's own view on these matters was that the inherence of our consciousness in an immortal soul can only be accepted on faith. The person that we are consists of a consciousness of present experiences and memories of past experiences. Whether that consciousness is born with the soul, stays with it throughout its life, and is retained by the soul after death, so that the soul does not run off and become someone else, is not something we can ever hope to know for sure. Locke speculated that the goodness of God would not permit that a soul that never willed or committed a wicked deed be sent to eternal damnation merely because by some accident it came to be imprinted with some criminal's memories. God will ensure, Locke hoped, that the same souls will keep the same memories, so that rewards and punishments are justly distributed. But the ties had been loosened, and alternatives, including the alternative that persons might just die and never be resurrected, were made possible. The only thing that allows us to suppose that the alternatives are not correct is faith, not a sure and certain knowledge of the way things must be.

These were profoundly unsettling consequences for many of Locke's more orthodox Christian readers.

ESSAY QUESTIONS AND RESEARCH PROJECTS

1. Outline Locke's theory of personal identity and then comment on whether phenomenon of false memory poses any problems for that theory. Supposing that someone can be convinced (say by a skillful police interrogator) that they remember committing a crime that they did not commit, would they in fact become guilty of that crime?
2. Outline Locke's theory of personal identity and then comment on whether the phenomenon of the transitivity of identity poses problems for that theory. We think that identity is transitive, that is, that if a=b and b=c then a=c. But, to cite a case popularized by Thomas Reid, we can imagine a man who as a boy was beaten for stealing cherries, who as a young adult seized a standard in battle, and who as an old man is a general. The general remembers seizing the standard, but not being beaten; the person who seized the standard remembers being beaten. Yet Locke's account seems to entail that while the general is the same person as the soldier who seized the standard, and the soldier is the same person as they boy who was beaten, the general is not the same person as the boy.
3. According to Locke's theory of personal identity, a person who cannot remember having committed a crime cannot be considered guilty of that crime, even if the person happens to now inhabit the body that performed that crime. This claim was challenged by Molyneux, who charged that we do take people who, say, get so drunk they cannot remember what they are doing and who then commit crimes to be responsible for those crimes. Locke's correspondence with Molyneux on this question is discussed by Henry Allison, "Locke's Theory of Personal Identity," in Ian Tipton, ed., *Locke on human understanding* (Oxford: Oxford University Press, 1977), P. Helm, "Did Locke Capitulate to Molyneux?" *Journal of the history of ideas* 42 (1981): 469-477, and Henry Allison and Nicholas Jolley, "Locke's Pyrrhic Victory," *Journal of the history of ideas* 42 (1981): 672-74. Survey this literature and comment on what it establishes.
4. Survey the recent philosophical literature on identity and personal identity. Have recent philosophers improved significantly on Locke's account and if so how?