

## Persons

I, a stranger and afraid  
In a world I never made.  
– A.E. Housman ([100], 109)

### 12.1 The raccoon's tale

In the fall of 1982, my department at Simon Fraser University mounted its fourth annual public-issues conference. The theme for that year was “Challenges to Science” and was widely advertised off-campus. The meetings attracted persons from many backgrounds. On the first day, a buffet lunch was served. Having taken a bit of tuna salad, my wife and I seated ourselves at a table with some strangers. The man on my left struck up a conversation.

“In my previous life, I was a raccoon”, he said.

Thinking this a bit of an odd icebreaker, I replied in what I assumed was the same spirit that the remark had been offered.

“I see. Do you feel a compulsion to wash your food in a mountain stream?”

I quickly discovered my mistake, however. The stranger had been in dead earnest. He firmly rejected my suggestion, and then persisted, not aggressively, but determinedly, in his claim.

“I was a raccoon before I was a person.”

Perhaps he said “human being” rather than “person”. My memory is not as precise as I would like on this particular point. I pressed him a bit.

“How do you know that? What makes you believe that you were a raccoon?”

The stranger was unable to offer any evidence beyond his own unshakable conviction that this was true. In some way, totally unanalyzable, and apparently not causing him any particular concern, he just ‘knew’ he had been a raccoon. At that, the topic had reached a dead

end and we turned to other, more usual, sorts of conversation.

In looking back on what was one of the most unusual exchanges of my life, I have had to ask myself several questions. What could it possibly mean for a person to have been a raccoon? Is such an idea even intelligible? Of course, if we try hard, we can imagine what it would be like to be 'housed' in a raccoon's body: instead of having a nose, one would have a snout; instead of hands, claws; etc. But this was not what that man had been claiming. He had not claimed that he – a person – had been housed in a raccoon's body; he claimed that he had *been* a raccoon. Putting aside the question why he might have thought such a thing, one must wonder what sort of theory of personhood we would have to adopt which would allow us even to imagine such a thing. For a raccoon to 'become' a person, for some 'thing', let us say  $x$ , to 'become' some later 'thing', let us say  $y$ , it is essential that something or other be preserved in the transformation: there has to be some 'important' connection between the earlier  $x$  and the later  $y$ . But what could this possibly be in the supposed case of a raccoon's becoming a person? According to the man who believed this of himself, it was not the *body* of the one which became the *body* of the other. Was it the mind? By his own admission, he had no memory of having been a raccoon. But how essential is memory for mind? Could the mind of a raccoon now be the mind of a man but without the man having a memory of having been a raccoon? If it was not mind, might it have been something else? Perhaps the *soul* of the raccoon became the soul of the man. But is this intelligible? What are souls? What counts for or against a soul's enduring and changing through time? In short, the claim provokes – and for our purposes serves to introduce – the cluster of problems concerning the analysis of personhood and of the identity through time of persons.

## 12.2 Persons and human beings

Every person I have ever known has been a human being. By "human being" I do not mean, as this term is sometimes used, "a decent, upright person", but rather I mean a living animal of the species *Homo sapiens*: a flesh-and-blood mammalian creature having a head, a torso, and typically two arms, two legs, etc., standing upright, breathing air, eating a variety of organic produce, etc.

While every person I have met, and expect to meet, is a human being, it is not at all clear that persons *must* be human beings or that all human beings are persons. At least for the moment we want to

leave it as an open question whether a person could have a nonhuman body (an animal body<sup>1</sup> or an electromechanical body<sup>2</sup> perhaps). Then, too, anencephalic infants (human beings born with no brain) who may be able to carry on some basic life processes are not conscious and have no prospects of consciousness. They are human beings, in that they have human (albeit defective) bodies, but it is arguable whether such grievously deficient, nonconscious human beings can be reasonably regarded as being *persons*.

One of the most difficult problems some persons have when they first approach these questions is to sort out the difference between the *legal* criteria for personhood and the *conceptual* criteria. The Law is a poor touchstone for deciding conceptual issues. The Law, in some jurisdictions, may rule, for example, that a fetus is a person. But although the Law may so rule, one can always ask, "Does this law comport with what our concept of personhood is? Do we have good philosophical grounds for accepting that law, or should we want to argue that it rests on a conceptual mistake and ought to be changed?" We are not logically, legally, or morally bound to accept the decisions of Law in constructing our own best concept of personhood. Ideally, the order of precedence ought to be the other way round: Law ought to try to capture the best thinking of the society; it ought to follow the best thinkers, not lead them.

Thus, even if 'the Law' (and of course 'the Law' is hardly monolithic, but varies from jurisdiction to jurisdiction, society to society) were to say "anyone is (legally) a person who satisfies the conditions *a*, *b*, and *c*", that would certainly not answer for us the question how we ought best to conceive of personhood. Even if in the eyes of the Law every human being were to be considered a person, that would not tell us whether from a considered philosophical point of view that was a warranted conclusion or not. We may be legally obliged to act in accord with the Law, but we surely do not have to believe or think in accord with the Law.

Then, too, the Law is nearly always reactive. It responds to needs

1. We are reminded of Kafka's Gregor Samsa (a gigantic insect), of Lucas's Chewbacca (a Wookiee), and of assorted Ewoks, werewolves, frog princes, etc.

2. Recall such fabulous characters as Pinocchio (more mechanical than electrical) and, of course, See-Threepio (also known as C-3PO) and RoboCop.

and disputes as they arise and become issues in the community. The Law seldom anticipates changing beliefs and thus does not plan in advance for them. But metaphysics, and philosophy in general, is different. Metaphysicians are free to speculate, and indeed considerably enjoy speculating, on situations which have not arisen – and indeed may never arise – in their attempts to refine our concepts. For metaphysicians, the Law may be a storehouse of case studies, a repository of much of traditional thought, but it can hardly serve as the arbiter of the cogency of a conceptual reconstruction.

The problems, then, to be addressed are these. Virtually every person is a human being; virtually every human being is a person. But *must* persons be human beings; *must* human beings be persons? Could a person have a nonhuman body? Might a human being be other than a person? In short, what is the *conceptual* connection between *being a person* on the one hand and *being a human being* on the other?

### 12.3 Why individuation and identity collapse in the case of persons

Anthony Quinton does philosophy in an admirably painstaking and systematic fashion. It is thus somewhat surprising to find, in reading *The Nature of Things* ([165]), that although he seems to be proceeding in a careful step-by-step fashion, examining first the problem of individuation of material objects and next the identity-through-time of material objects, when he comes to the subsequent discussion of persons, he skips over the question of the individuation of persons and proceeds immediately to the question of the identity of persons. Why the apparent omission? On the face of it, there is an entire chapter missing in his book, and yet – so far as I can tell – he offers not a single sentence of explanation as to why he departed from what looks to be the obvious and natural game plan. Might there be some reason why one would not treat the question of persons in a parallel manner to that already established for material objects, that is, by beginning with the question of the individuation of persons and then, in due course, graduating to the question of persons' identity-through-time?

I think there is a reason for *not* treating the question of persons in this two-step manner. And even if Quinton neglected to address the issue at all, I think we might do well to pause over it for a moment.

It is, of course, truisitic to say that persons are not 'just' material bodies. Persons may have material (in particular, human) bodies, but they also have properties and moral rights which no mere physical

body possesses. Persons can think and can act in ways that no 'mere' physical object, particularly a nonliving object, can remotely replicate. But these remarkable transcendent abilities are not what warrants leapfrogging over the question of personal individuation directly to the question of personal identity. The reason is slightly more concealed.

Material objecthood, i.e. being a material object, can be predicated of an existent thing on the basis of properties it instances, if not exactly all at one moment of time, then over a very short period of time. To count the objects in a room, for example, a procedure which requires that we individuate them, we will have to see which ones occupy space in the sense that they exclude other objects from the same space. We need this latter test to tell, for example, which are mere holographic images and which are 'real' physical (material) objects. But we do not need much of their history to individuate them; theoretically, a millisecond of endurance is adequate.<sup>3</sup> But there is no such equivalent determination possible for individuating persons.

Of course one could count the human bodies present. But while that is a good *practical* means, it is not entirely theoretically satisfactory. Some human bodies, even if alive, hardly are the bodies of persons. Human bodies born without brains, in which there is no consciousness whatsoever, can hardly be regarded as the bodies of *persons*. And again, it is theoretically possible that a person should have other than a human body. In short, at the very least, at the outset of our examining the question of the individuation and identity of persons we do not want to prejudice the issue by assuming that persons *must* be identified with living human beings. Perhaps at the end of our researches we may want to assert such a thesis. But if we do, then such a thesis is something to be argued for, not assumed from the outset. It ought, that is, if it is to be promoted, to be argued for as a conclusion of an argument and not assumed as a premise.

In skipping over the question of the individuation of persons, directly to the question of their identity, we do so because we already have an eye on our eventual conclusions. To be a person is essentially,

---

3. That this is so depends very much on certain physical facts characteristic of this particular world. In a Waismanesque world, recall (p. 301), it would be necessary to know something of the remote history, viz. the details of its manufacture, of a seeming single chair to know whether it was in fact one chair or a pair.

among other things, to be the sort of thing which does (or, in the case of newborns, will) have memories. But to have memories requires that the person be extended in time. There logically cannot be short-lived persons (e.g. having a duration of a millisecond) in the way, for example, there can be short-lived physical objects, e.g. muons whose lifetime is of the order of two-millionths of a second. It is of the essence of being a person that one have a history of experiences.

#### 12.4 Is there a self?

In *A Treatise of Human Nature*, David Hume begins his discussion of personal identity with what, at first, seems to be nothing more than a casual, innocuous, recounting of a common belief among philosophers about our direct acquaintance with our selves.<sup>4</sup> (Although Hume restricts his discussion to the beliefs of fellow philosophers, he might just as well have spoken of vast numbers of persons educated and living in Western culture.) “There are some philosophers, who imagine we are every moment intimately conscious of what we call our SELF; that we feel its existence and its continuance in existence; and are certain, beyond the evidence of a demonstration, both of its perfect identity and simplicity” ([101], book I, part IV, section VI, 251). But Hume does not broach this topic of self to lend his assent to the commonly held view; he raises this issue of self in order to probe it and, eventually, to reject the common conception. In one of the most celebrated passages in all of philosophy, he shortly continues:

For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never catch *myself* at any time without a perception, and never can observe any thing but the perception. When my perceptions are remov'd for any time, as by sound sleep; so long am I insensible of *myself*, and may truly be said not to exist. And were all my perceptions remov'd by death, and cou'd I neither think, nor feel, nor see, nor love, nor hate after the dissolution of my body, I shou'd be entirely annihilated, nor

---

4. This way of putting the point is not Hume's, but a modern reconstruction using terminology, viz. the term “acquaintance”, which has been borrowed from Russell. See p. 309 above.

do I conceive what is farther requisite to make me a perfect non-entity. If any one upon serious and unprejudic'd reflexion, thinks he has a different notion of *himself*, I must confess I can no longer reason with him. All I can allow him is, that he may well be in the right as well as I, and that we are essentially different in this particular. He may, perhaps, perceive something simple and continu'd, which he calls *himself*; tho' I am certain there is no such principle [elemental thing] in me. ([101], 252)

If this passage is read superficially, it gives the appearance of being self-refuting, for Hume writes “*I* enter ...”, “*I* call ...”, “*I* conceive ...”, etc. Do not his very own words betray the impossibility of his maintaining what he claims, viz. that he cannot find *himself*? Is not saying, as Hume does, “*I* am insensible of *myself*”, as self-refuting as saying, “*I* do not exist”?

Once again (recall our earlier discussion, p. 171), we find a philosopher denying that something exists which is thought to be familiar to great numbers of other persons. And again, just as in other cases, we find in this instance that the philosopher is denying one thing only to assert another.

Hume is not, of course, denying that he exists. He is perfectly comfortable speaking of himself and using the personal pronoun “*I*” of himself. What he is denying is that ‘self’ is anything ‘given’ in perception. And what he is offering on the positive side is the thesis that self is “a bundle or collection of different perceptions, which succeed each other with inconceivable rapidity, and are in a perpetual flux and movement” ([101], 252). In short, Hume’s theory is that self is nothing more or less than ‘a bundle of perceptions’, or – to use a more modern vocabulary, and one not restricted solely to perceptions – self is a ‘stream of consciousness’.<sup>5</sup> Sometimes this theory is also called the ‘no-ownership’ theory, since it argues that there is no self which owns or possesses the succession of items in that stream of consciousness: the items follow one another, as Hume says, “with inconceivable rapidity”, but they are not ‘in’ or ‘of’ a self.<sup>6</sup>

---

5. More exactly, he maintains that self is a punctuated, or interrupted, stream, since we all have periods of dreamless sleep.

6. Some persons have argued that Hume’s experiment is naive in that absolutely constant things are, by their very nature, imperceptible and that what Hume was trying to perceive is something which would amount to a

Some philosophers have agreed with Hume. Richard Taylor, for one, goes even further than Hume ventured.

One imagines that he is deeply, perpetually, unavoidably aware of something he calls “I” or “me.” The philosopher then baptizes this thing his *self* or perhaps his *mind*, and the theologian calls it his *soul*. It is, in any case, something that is at the very heart of things, the very center of reality, that about which the heavens and firmament revolve. But should you not feel embarrassment to talk in such a way, or even to play with such thoughts? As soon as you begin to try saying anything whatever about this inner self, this central reality, you find that you can say nothing at all. It seems to elude all description. All you can do, apparently, is refer to it; you can never say what is referred to, except by multiplying synonyms — as if the piling of names upon names would somehow guarantee the reality of the thing named! But as soon as even the least description is attempted, you find that what is described is indistinguishable from absolute nothingness. ([204], 122)

Taylor knows full well the common conception of self which he is bucking. And he is as eloquent in presenting the view he wishes to refute as he has been in denying it. Taylor gives expression to the commonly held, opposing, view this way:

There seem to be two realities – myself and all the rest. By “all the rest” is meant the whole of creation except me. ... This rest, this everything else, all that is outside, other, is perpetually

---

*constant* element in perception. The objection continues by arguing that perception operates, essentially, by taking cognizance of *differences*. An undifferentiated, constant element of perception, coextensive\* in time with one’s entire existence, would be an impossibility since it would lack a contrast.

Such an objection relies on certain empirical claims about perception, claims which are exceedingly difficult to test. We know that we become desensitized to long-lasting stimuli: for example, we grow inured to a constant aroma, being unable to smell it at all after a long exposure. But still, such data fall short of proving that it is impossible to detect a truly constant element in perception. The claim is more metaphysical than empirical and, even at that, not particularly self-commending or self-evident. I mention this debate, but side with neither party to it.



changing, never two moments the same. But at the heart of it all, at that point which is the metaphysical center of my reality, is that self, that which is not something “else” – and it does *not* change, or at least does not become something *else*. It remains one and the same, throughout all the changes it undergoes, preserving its identity through an ever elapsing and growing time. Except for this – that it does finally suffer that calamitous change, which is its own extinction! And that is a pretty awesome thought, a dreadful thought, a cosmic insult. ([204], 122)

Taylor argues (not at all well or convincingly, in my opinion) against this latter view of *self*. And he concludes:

We wanted something [i.e. a self] to present as an ultimate reality, to contrast with everything else, and we found total, perfect nothingness! It isn't there. Imagination creates it. Intellect distinguishes it. Metaphysics builds intellectual fortresses upon it. Religion guarantees its salvation – always, of course, on certain terms – and promises to push back the nothingness that approaches it. And all the while, it is itself the most perfect specimen of nothingness! One does indeed feel like a child discovered making a face at himself in the mirror. One wants somehow to cover up what was going on, embarrassed at his own ridiculousness. ([204], 123)

Taylor's counsel, then, for those afflicted with the Dread of Death: The self cannot die, for there is no self.

Hume's and Taylor's extreme ideas about self are heady, perhaps alarming, and for some persons, even frightening.<sup>7</sup> But they are also important, if for no other reason than to cause us to shake off our complacent, comfortable misconception that there is any *universal* idea of selfhood. For theirs is but one of a bewildering array of quite different notions of what *self* might be.

Visit any well-stocked library and look at the number of books catalogued under the subject heading “self”. (And look, too, at the number dealing with “death”.) The figures are staggering, and writers from an

---

7. They are also threatening if read by someone reared in a religious tradition where *self* (or soul) is a central concept and where children have been taught not to question church dogma.

enormous number of fields all contribute to the froth: philosophers, to be sure; but also psychologists (of every imaginable stripe, Freudian, Jungian, Existentialist, Experimentalist, Behaviorist, etc.); sociologists; anthropologists; educators; criminologists; novelists; essayists; historians; etc.<sup>8</sup> There is, in fact, a veritable industry given over to generating an endless supply of articles, novels, and learned books on the topic of *self*. Our collective curiosity on this topic (like that on sex and diet) seems limitless.

We will confine our attention to the major philosophical theories of personal identity.

### 12.5 The principal contemporary theories of personhood

For a person *a*-at- $T_1$  to be identified with a person *b*-at- $T_2$ , there must be some thing, or set of features, which unifies the two, which accounts for their being two stages of one and the same person. The principal theories are these:

- The unifying principle is soul (self, or mind).
- The unifying principle is physical body (usually, if not invariably, a human body).
- The unifying principle is similarity between successive bundles of sensations.
- The unifying principle is personality and memory.
- The unifying principle is an amalgam of various of the preceding.

Although I have never taken a poll, my own educated guess, arising from my having been brought up and exposed to much the same sort of culture as everyone around me, is that the theory that it is soul (or self) which accounts for a person's identity is the most widely held one of the lot. It has, however, steadily, and perhaps at a quickening pace in modern times, been losing some of its original religious trappings. Many persons who are not religious still cling to a concept of soul not terribly unlike that historically promoted by Christianity. Many non-Christians retain the belief that the soul is, in some fashion, not a physical thing, but a supernatural sort of entity. Where their

---

8. I will ignore all the execrable "self-help" books written by an army of poseurs and dilettantes whose scientific credentials are often vanishingly close to nil.

notion departs from traditional religious views is in their abandoning the further belief that souls endure beyond bodily death. They have come to adopt a 'secular' concept of soul. Soul is posited to be what makes one person different from another, while at the same time souls are believed, somehow, to be dependent for their existence upon the existence of a living human body. All in all, soul appears as a mysterious 'I know not what' which plays the theoretical role of providing the basis for personal identity. Soul, while not physical, is what is conceived of as being what is essential in a person. The positing of soul as a solution to the problem of personal identity is a positive theory analogous to the positing of substance as a solution to the problems of individuating and re-identifying physical objects. And like the theories of material substance, it encounters similar sorts of metaphysical and epistemological difficulties.

Some persons, we all know well, have an unshakable conviction that souls exist. They are as sure of the existence of souls as they are of tables and chairs. Other persons are less sure; and some persons, of course, are convinced that souls do not exist. Souls (if they exist) are not publicly perceivable things. We cannot prove or demonstrate the existence of souls by holding them up for public display, or by pointing to one, or even by directing persons to introspect and thereby to discover their own souls. Hume, as we saw a moment ago, tried the exercise and reported abject failure. So did Taylor. This is not to say that everyone who tries the exercise must fail. Hume knew that some persons might try the experiment and come to believe that they had succeeded in finding their souls. But Hume's, Taylor's, and many other persons' reported failures do tell us that searching for one's soul by introspection is not a test which yields anything like universal agreement. And it is not unreasonable to hypothesize that persons who do report success in administering the test to themselves are persons who were antecedently disposed strongly to believe that they had souls. From a methodological point of view, successful reports of the experiment must be regarded as tentative at best and, perhaps, suspect as well.

But my purpose here is not so much to argue against the existence of souls as it is to point out that their existence is problematic. And that their existence is problematic is all that is needed to render souls inappropriate platforms on which to erect solutions to the problem of personal identity.

There are, to be sure, some exceedingly troubling cases in which we might be terribly unclear whether personal identity has been preserved

or not. (We shall turn to some of these cases presently.) But in both ordinary and troublesome cases, we never actually proceed by trying to detect the person's *soul* and asking ourselves whether it is, or is not, identical with some earlier (or later) soul. The fact is that, whatever might be our final opinion about the existence of souls, we never actually invoke the concept of soul in our day-to-day re-identification of other persons or even, for that matter, of ourselves. You may catch a glimpse of someone on the street who looks like a long-lost friend. "Could that be Jim?" you ask yourself. In an impulsive mood, you shout, "Jim!" He turns, stares blankly for several embarrassing seconds, and then flashes a familiar grin of recognition. Identification has been made. And neither you nor Jim has examined the other's soul.

If one adopts the theory that sameness of soul confers personal identity, then one can make sense of the claim of the man who believed that he had in a former life been a raccoon. He had been a raccoon if the soul he now has formerly had been the soul of a raccoon.<sup>9</sup> But while we may, by adopting this theory of soul, be able to attribute a *meaning* to his claim, we will not have succeeded in making that claim *rational*. To make such a claim rational, we would have to have some account of how it is possible to *know* such a thing or to have good evidence for it. And inasmuch as the very existence of soul seems so problematic, the belief that there could be objectively valid criteria for re-identifying souls seems utterly forlorn. The price of adopting a theory of souls as personal re-identifier is the abandoning of rational grounds for making identifications.

From an epistemological point of view, souls are idle: they play no role in our day-to-day identification of other persons. From a metaphysical point of view, positing souls as the principle of personal identity is, as was the case with material substance, regressive. It simply displaces the problem of identity, but leaves it otherwise unresolved. By arguing that person *a* is identical with person *b* if and only if the soul of *a* is identical to the soul of *b*, we have merely deferred the question, but not answered it. For now we must ask whether the soul of *a* is identical to the soul of *b*. And whatever way

---

9. This is of course also to assume that raccoons' souls are not so different from persons' souls that the one could not become the other. I shall not pursue this baroque question whether raccoons' souls, horses' souls, turtles' souls, etc. are interchangeable or not. As you might suspect, I regard the exercise as ludicrous.

we might go about answering this latter question, we might well have pursued directly, that is, without having interposed the superfluous intermediary concept of *soul* in our attempt to answer the original question. In short, introducing the concept of *soul* does no useful work in helping us solve our real problems: it merely retards the progress toward a solution.

Positing soul is, thus, not going to be of much help in solving either the epistemological or the metaphysical problems of personal identity. We must seek another identifier of persons.

The theory that it is the human body which is the identifier is considerably more promising: it is economical in the sense that it assimilates the problem of personal identity to that of the identity-through-time of a particular material object, viz. a person's own body, and it invokes no hidden or exotic substances. Moreover, it is, after all, clearly the criterion we daily use in identifying other persons. How did you recognize Jim, and he, you? By noticing certain familiar physical features in the appearance that one presents to the rest of the world. (Human beings have an uncanny ability to recognize extremely subtly different features of physiognomies.)

But John Locke, recall (section 6.4 above, pp. 108ff.), argued strenuously against using the practical criterion of bodily identity as the theoretical criterion of personal identity. Although the body may be used as a surrogate criterion, it was not to be regarded as ultimately satisfactory: "... should the soul of a prince, carrying with it the consciousness of the prince's past life, enter and inform the body of a cobbler as soon as deserted by his own soul, everyone sees he would be the same person with the prince, accountable only for the prince's actions" ([124], book II, chap. XXVII, §15). The talk here of the transference of soul from the prince's body to the cobbler's is incidental; the essential aspect is not the transference of soul – Locke is very emphatic on this point, reiterating it several times – but the element of the transference of *consciousness*. It is the transference of consciousness, alone, which makes for the transfer of the prince to the cobbler's body. The identity of persons is grounded in consciousness, not soul. Indeed, Locke argues at some length that a person could successively have different bodies, that a person could take turns sharing a body with another, that a person could have different souls and might even now have the soul of some former person, but that none of this would affect that person's identity. For a person's identity is not a matter either of body or of soul, but strictly of consciousness.

For Locke, the element of consciousness which played the crucial role in personal identity was memory. To the extent that a person has the memories of an earlier person, to that extent he/she may be identified with that earlier person. In one of his not infrequent heroically convoluted sentences, Locke writes:

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 write this now, that saw 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 that write this am the same *myself* now whilst I write (whether I consist of all the same substance, material or immaterial, or no) that I was yesterday. ([124], book II, chap. XXVII, §16)

More simply: "If I had memories of seeing Noah's Ark and the worldwide flood as well as memories of the Thames overflowing last winter which were as compelling as the perceptions I am now having of writing this passage, then I could not doubt that *I* did indeed see the Ark and Noah's flood and that *I* saw the Thames overflow last winter."

Joseph Butler (1692-1752), writing more than thirty years after Locke's death, challenged Locke's theory by arguing that Locke had got the order of logical priority reversed, that it was personal identity which accounts for memory, and not the other way around:

But though consciousness of what is past does thus ascertain our personal identity to ourselves, yet to say, that it makes personal identity, or is necessary to our being the same persons, is to say, that a person has not existed a single moment, nor done one action, but what he can remember; indeed none but what he reflects upon. And should one really think it self-evident, that consciousness of personal identity presupposes, and therefore cannot constitute, personal identity; any more than knowledge, in any other case, can constitute truth, which it presupposes. ([39], 298)

Memory, for Butler, is *evidence* for personal identity, but does not itself constitute personal identity. For memory to be evidence of per-

sonal identity, personal identity must itself exist *independently* of the evidence for it. He continues, giving voice to an intuition which is antithetical to Locke's own:

... though present consciousness of what we at present do and feel is necessary to our being the persons we now are; yet present consciousness of past actions or feelings is not necessary to our being the same persons who performed those actions, or had those feelings. ([39], 298)

That Locke and Butler disagree, and that their disagreement stems from totally different prephilosophical beliefs about the centrality of the concept of *memory* to the concept of *person*, is apparent.<sup>10</sup> But is this the end of the matter? Must this debate simply be regarded as a clash of intuitions, and must it be left at that?

In the very last paragraph of his essay, almost as an afterthought, Butler raises an issue which has come to be seen as essential in tackling these problems. For Butler reminds us that memories can be *mistaken*. And although Butler, himself, does not particularly pursue this problem, it really does pose a crucial difficulty for Locke's theory.

According to Locke, personal identity is constituted by memory. But what if one's memory is mistaken? What if someone is convinced that he recalls something, but his report is about an event at which he could not possibly have been present? (This need not be regarded as a pathological condition. All of us have mistaken memories about some things. Sometimes we might believe that a dream was a 'real' memory. And modern empirical research has shown just how much eyewitness accounts of 'one and the same event', even among persons

---

10. Compare Butler with Hume (publishing three years later [1739]): "Who can tell me, for instance, what were his thoughts and actions on the first of *January* 1715, the 11th of *March* 1719, and the 3d of *August* 1733? ... Will he affirm, because he has entirely forgot the incidents of these days, that the present self is not the same person with the self of that time; and by that means overturn all the most establish'd notions of personal identity? In this view, therefore, memory does not so much *produce* as *discover* personal identity. ... 'Twill be incumbent on those, who affirm that memory produces entirely our personal identity, to give reason why we can thus extend our identity beyond our memory" ([101], book I, part IV, sect. VI, 262).

who are trying their level best to be scrupulously honest, can differ markedly.)

Locke's own example is of someone (himself, of all people) possibly recalling having witnessed the biblical Flood. Locke certainly is consistent: he allows that any person who has such memories was present at the Flood.

Most other persons, from Butler onwards, are far less likely to be quite so liberal. Were any of us to meet someone who claimed to have been present at the Flood, most of us, I am sure, would be skeptical in the extreme, probably believing in the first instance that the person making the claim was mentally ill or suffering some sort of delusion. Does this mean that one must, then, adopt Butler's theory, that personal identity is the basis for memory, and reject Locke's, that memory is the basis for personal identity?

That there is a problem in Locke's theory does not, of course, mean that Butler's opposing theory is correct. Butler's would be the preferred theory only if these two theories were the only ones possible. But they are not. And indeed, what I want to suggest is that what is needed is not the wholesale rejection of Locke's theory, but a repair.

I am convinced, like Locke, that memory does play a central role in personal identity. But the role cannot be as simple and direct as Locke imagined. For Locke's insights can be invoked only for correct memory (or veridical memory, as it is sometimes called), and not for mistaken (or falsidical) memory.

But what is the test of veridical memory? We have already explored this question (in section 8.10, pp. 220ff.). There I argued that one way to test memories is to compare one's own 'seeming' memories with those of other persons. If they agree, then one has good *prima facie* evidence of the correctness of one's own memories. But what if others' memories do not bear out one's own, or what if other persons were not witnesses to the event you believe you recall, or what if – even more extremely – your memory is of an event predating the birth of anyone alive today? How then shall it be tested? As I argued earlier: by consulting the testimony provided by physical facts. Ultimately the reliability of memory, and our ability to sort out veridical from falsidical memories, at some point must rely on the evidence of the physical world.

If this were a world where persons never had bodies, where they were just thinking things, then one might want to argue that insofar as there would be no way to distinguish veridical from falsidical memories, there would not be such a distinction, and that having a



memory, any memory however wild or bizarre, would then be a memory of one's own personal history. But such a world is vastly different from this world.

In the last few decades, there has been a marked shift in much writing about personal identity. Whereas in centuries past philosophers were disposed to ground personal identity in souls, in substance, and other empirically problematical entities, many recent philosophers are disposed to seat personal identity, as did Locke, in memory and, many would add, personality. But they do not rest there. There is more to the concept of personal identity. Not just any memory, or seeming memory, will do. It must be authentic, or veridical, memory. And for memory to be veridical, we normally require that the body (of the person whose memory it is) was present at the remembered event. In short, although the body is not the identifier itself, it plays a crucial role in determining the authenticity of the identifier, viz. memory and personality.

Persons are essentially identified by their personalities and by their authentic (veridical) memories. But for memories to be authentic, the person must be embodied, i.e. the test of cogency of memories depends on causal links in the physical world. (This is not, of course, to argue that memories are not themselves physical entities. They may be. As we saw in chapter 10, memories perhaps are states of our central nervous systems. But the theory of personal identity being proposed here does not require any particular decision in that latter case. All that is required is that memories – whatever their ontological fate, whether regarded as themselves physical states or not – be *testable* by the evidence furnished by physical states.)

Interestingly, another consideration, from quite another direction, also favors the theory that persons must be embodied. Recall our earlier discussion (p. 131) of Plato's allegory of the cave. There I argued that were persons not to be embodied, they could not tell 'themselves' apart from 'other persons', there could be no concept of *personal identity*.

What is emerging is a theory of personal identity which to a certain degree mirrors that of physical object identity. What confers identity is not the endurance of a mental or spiritual substance, but a succession of 'person-stages' unified, or integrated, by certain sorts of relationships they bear to one another.

Hume had grasped a fragment of this modern account. He, too, conceived of personal identity as a series, a succession, of stages. When

he looked for the 'organizing principle' he singled out a pair of relations which together he supposed conferred the identity: resemblance and causation. But these relations will not do. The role of causation is overstated. Some items in the stream of consciousness *may* causally bring about their successors, but an equal if not greater number of these episodes are induced by external stimuli. And resemblance fares little better.

For what is the memory but a faculty, by which we raise up images of past perceptions? And as an image necessarily resembles its object, must not the frequent placing of these resembling perceptions in the chain of thought, convey the imagination more easily from one link to another, and make the whole seem like the continuance of one object? In this particular, then, the memory not only discovers the identity, but also contributes to its production, by producing the relation of resemblance among the perceptions. ([101], book I, part IV, sect. VI, 260-1)

In commenting on this passage, Quinton offers this counterexample: "Suppose two men, A and B, take turns looking through a keyhole at moments 1 and 2. Then experiences A1 and B2 will probably be more alike than A1 and A2 or B1 and B2" ([165], 320). Quinton seems to be suggesting that Hume had argued that the relation of similarity was supposed to obtain between a person's *successive* perceptions. But that is not what Hume claimed (at least it is not what I take him to be writing in the passage above). Rather Hume claimed that in the stream of consciousness there will recur similar episodes (not necessarily successive to one another), and it is these recurring and similar episodes which contribute to the appearance of a unity.

The precise interpretation is a quibble, however. For it is clear that resemblance among the episodes in the stream of consciousness, whether those episodes are neighboring ones or remote from one another, will not unify the series. The point is that the series will be unified if the episodes are those of one person, and it will not be unified if the episodes – however much alike, regardless whether immediate neighbors or remote in time from one another – are those of different persons. You and I might at virtually the same time have qualitatively identical perceptions of a scene, and yet your perception is yours and mine is mine. And nothing intrinsic to our perceptions –

certainly not any relation of similarity (or dissimilarity) obtaining between the two – accounts for the one's being mine, and the other's being yours.

Any viable theory of personal identity is going to have to accommodate a remarkable variety of data.

- Introspection does not seem to reveal any organizing 'principle' (e.g. soul).
- Different persons, *a* and *b*, may have experiences,  $e_a$  and  $e_b$ , which are more alike one another than those experiences are like other experiences of *a* and of *b* respectively.
- Personality and veridical memory seem to play a crucial role in determining a person's identity.
- That a memory is veridical can be objectively established only if a person is (or at least has been) embodied.
- If persons are not embodied, then there is no objective test for distinguishing between self, hallucinatory 'other persons', and genuine other persons. Without a body, the distinction between 'self' and 'other' collapses.
- We virtually always use, as our practical criterion of personal identity (particularly that of other persons), the bodily criterion.
- It is perfectly intelligible to describe two persons swapping bodies. Few of us have any difficulty imagining ourselves having (being housed in?) a different body. The body, then, is not the ultimate, or sole, criterion of personal identity.
- Memories are constantly being lost. Some of what I did remember yesterday, and some of what I could have remembered yesterday, I cannot recall today. Some memories are very long-lasting; but others fall away. Memories may be likened to the physical parts of an object which are from time to time discarded and replaced by others. But whereas physical parts are often replaced by qualitatively identical parts, the greater part of our store of memories often changes markedly over a period of years.
- Both memories and personality traits are *dispositional*. Each of us is capable of recalling vastly greater numbers of events than any of us actually recalls at any one moment. Each of us acts and reacts to situations in idiosyncratic (personal) ways, but only one, or very few, of these will be exhibited at any one time. That is, the bulk of one's own memories and personality lies dormant, metaphorically speaking, ready to be activated,

but generally is not manifest. The only respectable theories we have of the nature of (instanced) dispositions are theories which would make of them (instanced) properties of some persisting, or enduring, thing. In the case of memories and personality, this would mean that memories and personality must be seated in 'something' which endures. If souls are not to be invoked for this latter role, then the most plausible alternative candidate is an enduring physical body. And of physical bodies, the human central nervous system is, by far, the most attractive and likely candidate to fill the role.

The term "personality" bears intimate etymological ties to "personhood". That it does attests to some of the metaphysics built into our language. But for the moment, I do not wish to invoke this question-begging aspect of "personality". I want to use the term "personality" in a more neutral way, without presupposing that personality is conceptually tied to the concept of personhood. Let us, then, for a while, suspend our recognition of the verbal link, and let us conceive of persons' personalities as the characteristic ways they react to situations. Obviously, one's character (one's moral and ethical dispositions) is part of one's personality, but "character" and "personality" are not synonyms. It may be part of your personality to like piano sonority, but we would not be much inclined to regard that liking as part of your character. In any event, for a moment, let's use "personality" as an abbreviation for "characteristic behavior".

Imagine a world where the personalities (as just defined) and the memories of persons could be swapped between bodies. Assume, too, that such swapping occurred universally, quite naturally, i.e. as an operation of Nature itself, to all persons, every day, worldwide at local noon.

What sort of social practices would a society have to institute to cope with such a phenomenon? Suppose a woman left her house at 8:00 AM. At noon her personality and memories are suddenly switched to another body. (We'll assume that body switching is always from male to male and female to female, youngster to youngster, and senior to senior.) At 4:00 PM she sets out for home. Which house should she return to: the house she recalls leaving that morning, or the house from which her body departed that morning? (Since she has no memory of the house from which her body would have departed, we would have to assume that were the latter alternative to be the adopted one, human bodies would have to be tattooed with their home addresses.)

I think *most* of us would be strongly inclined to opt for the first alternative, arguing that personhood is carried by personality and by memories, not by body. Were we to adopt this suggestion, then in the world just described, the practical criterion of identifying persons by their bodies would be fairly useless (it would never work over a time interval of twenty-four hours). Instead we would probably set up some manner of greeting one another whereby we would exchange names or other information which would uniquely identify ourselves to one another.

Of course, there is no *necessity* – either physical or logical – to adopt the practice just described. A society theoretically could adopt the second practice, i.e. of identifying persons by their bodies, not by their personalities and memories. In such a society, when husband and wife greeted one another each evening, the bodies returning home would be those pictured in the photograph on the mantle, but the interests, memories, and personality of each person would be entirely unfamiliar to the other. Such a practice might work. (And some among you might even be intrigued by the prospects, believing that it would relieve the humdrum in ordinary life and make the principal causes of marital breakdown disappear at a stroke.) But the fact is that the concept of personhood which would be implicit in this latter practice is not the concept we use. In this world, in our circumstances, we conceive of personhood as seated, not in body, but in personality and memory.

How can we tie this all together? I suggest in this way. Our concept of *person* is built on the requirement that identity of persons is secured through (genuine) memory and personality. (We will explore each of these requirements further in the subsequent two case studies.) But personality and genuine memory *presuppose* embodiment. Persons must be embodied in order to individuate them and in order to distinguish genuine memory from hallucination and delusion. But this is not to say that a person must have exactly one body throughout his/her lifetime. The requirement of embodiment is satisfiable by a person's having a succession of bodies. What is essential is memory and personality, but that memory and personality must be embodied.

This criterion conceals several imprecisions. From day to day we might recall the greater part of what we could recall the previous day. Today's memories are fairly similar to those of yesterday. But this similarity of memories, from day to day, does not hold for days much further separated. When relatives tell me of things I did when a toddler, I have no memory at all of having done them. I can, today, recall

nothing whatsoever of my fourth birthday. There is certainly a spatio-temporal continuity between that youngster's body and my own: I still bear the physical scars of some of his mishaps. But am I to be regarded as the *same person* as that four-year-old of yesteryear? I am sure intuitions will diverge significantly on how to answer this latter question. And even among persons who will want to insist that the adult is the 'same person' as the former child, there may well be a debate as to whether their grounds for saying this depend on their making the identification on the basis of the spatiotemporal continuity of the human body or whether they depend on there having been a day-to-day (but not year-to-year) similarity of memories and personality.

There will inevitably be a strong temptation to assimilate this present conundrum to that examined earlier, viz. Hobbes's version of the problem of the ship of Theseus (pp. 347ff.). In that earlier instance, there were two competing criteria of identity: the spatiotemporal one and the compositional one (i.e. the criterion of re-identification by material parts). I argued that the former is primary, but in situations where it is inapplicable, then it is appropriate to fall back upon the latter criterion. Were we to apply that sort of reasoning to the present case, we might argue that inasmuch as the adult cannot remember having been the child (i.e. where the criterion of continuity of memory and of personality is not satisfied), one may fall back upon the strictly physical criterion of the spatiotemporal continuity of the human body.

But the analogy is not nearly so simple. There are profound implications in identifying persons. Locke warned of the danger in our falling back upon the bodily criterion of personal identity. For instance, he considered it an abomination to punish a person for misdeeds of which he had no memory:

... if it is possible for the same man [i.e. human being] to have distinct incommunicable consciousness at different times, it is past doubt the same man would at different times make different persons; which, we see, is the sense of mankind in the solemnest declaration of their opinions, human laws not punishing the *mad man* for the *sober man's* [i.e. the normal man's] actions, nor the *sober man* for what the *mad man* did ... . ([124], book II, chap. XXVII, §20, 287-8)

There is something deeply troubling in the prospect of punishing a

person for a crime of which he has utterly no memory.<sup>11</sup> The spectacle of a person suffering a heavy fine, or sitting in a jail cell, for having done something which he cannot recall at all strikes many of us as a miscarriage of justice. There can be no contrition by a person who has no memory of having committed an offense; there can be no personal guilt.

Thus, the decision to use the bodily criterion of personal identity as a fallback option must not be undertaken lightly. It is not a mere convenience. Adopting the bodily criterion of personal identity in cases

11. There are exceptions. Locke, himself, allowed for the case of punishing a person who committed an offense while intoxicated. But Locke was troubled over the rationale, or justification, for this practice. He believed that punishment was permissible because we “cannot distinguish certainly what is real, what counterfeit [in cases of drunkenness or sleepwalking]; and so the ignorance in drunkenness or sleep is not admitted as a plea” ([124], book II, chap. XXVII, §22). This is certainly a wretched justification for our legal practice. (I will ignore the case of the sleepwalker and concentrate solely on the case of the offense committed while intoxicated.) Locke argues that since there is no way to prove that a person who claims ignorance of his drunken actions is telling the truth, it is permissible to punish him. This seems to have turned the principle of ‘innocent until proved guilty’ on its head, placing the burden of having to prove himself innocent on the accused instead of placing the burden of having to prove the accused guilty on the prosecution. Locke defends this violation of the principle on the grounds that on “the Great Day” (i.e. Judgment Day), it will all be put right. (But see section 12.8 in this chapter.)

There is, however, a far more reasonable justification – stemming from Aristotle ([11], book III, 1113b29-1114a3) – for punishing a person for his offenses while drunk even if he cannot now recall committing those offenses. When persons *choose* to drink alcohol, they do so in full knowledge that they might commit an offense and might lose the memory of having done so. The subsequent loss of memory is not something which just randomly happens to befall the drinker; quite the contrary, he chose to do something (drink alcohol) which he knew might very well blot out memory. It is this aspect of the affair – knowingly taking a drug which might precipitate one’s committing an offense and which also might blot out one’s memory – which justifies our subsequently holding the person responsible for his misdeed. If we did not have such a practice, then if there were a memory-erasing drug, anyone could absolve himself of guilt by taking that drug after having committed a crime and wiping clean his memory of the offense. I think few of us would be inclined to regard his after-the-fact self-induced loss of memory as warranting the dropping of proceedings against him.

where the criterion of continuity of memory and personality is inapplicable has profound *legal* implications.<sup>12</sup>

In any event, it is not my purpose to pursue the nuances of legal reasoning. I am prepared to leave the debate at this point, to return to an exploration of some of the other implications of adopting the sort of analysis which has been evolving here. In the course of the ensuing case studies, I will try to expose something more of the vagueness of the concept of identity and will suggest that the idea that there is, or can be, some precise notion as to the complete set of essential ingredients in the concept of personal identity is impossible to realize.

## 12.6 Case study: Tim's plaint

One of my closest friends in graduate school was a history buff. Tim (not his real name) felt himself a misfit in the then-current (viz. the twentieth) century. He loathed the pace of life, the congestion, and especially the suffering and devastation wrought by modern warfare. Often, in perfect seriousness, he would lament to me that he had been born in the wrong century. Tim sincerely wished that he had been born in and had lived his entire life in the seventeenth century, whose life-style he regarded as being far better suited to his own particular temperament, needs, and attitudes.<sup>13</sup>

On those occasions (in 1963), when Tim would begin to express such unrealizable desires, I was fully prepared to enter with him into his fantasy and to 'play' by his rules. At that time my usual response was to remind him of all the benefits which living in the twentieth century bestowed and of all the advantages persons living in the seventeenth century did without. I reminded him that he had been

---

12. It is interesting that we probably would feel rather more sanguine about adopting the bodily criterion where the consequence would be the bestowing of a good rather than the exacting of a punishment. Those of us who might protest the punishing of a man who had no memory of having committed an offense might be far less moved to complain in the case of an adult's inheriting a legacy even though he has lost his memory of having earlier been the child whom the legator had originally designated as being the recipient.

13. Tim wanted to have been born earlier, not to have never been born at all. Bernard Williams reports that there is an 'old Jewish reply' to the latter request, i.e. to have never been born. It is: "How many are so lucky? Not one in ten thousand" ([212], 232).



seriously ill a few years earlier and that modern pharmaceuticals had saved his life. Persons in the seventeenth century who had such illnesses never recovered. And I reminded him, too, that persons living in the seventeenth century knew nothing of the music of Beethoven, Schubert, Brahms, Dvorak, Verdi, Puccini, Prokofiev, Weill, Gershwin, and Brel. Persons living in the seventeenth century were ignorant of the writings of Hemingway, Tolstoy, Dostoevsky, and Dickens. They would never have heard the voices of Caruso, Gigli, Galli-Curci, Robeson, Bjoerling, Milanov, Piaf, Jolson, and Lenya. They would never have heard performances by Heifetz, Horowitz, and Gould. They would never have seen the films of Bergmann, of Welles, and of Hitchcock. They would never have savored the wit of William Gilbert, Ogden Nash, and Lewis Carroll. They would never have seen the sculptures and paintings of Rodin, Picasso, and Miro. And the list went on and on.

That was half a lifetime ago. As you can see, my argument, then, focused on selling the triumphs of our own times, and it involved a recitation of a variety of highlights of the previous one hundred years or so. In the intervening decades, however, I have often reflected on Tim's plaint, and I have come to have a totally different perspective.

Earlier (in section 8.11) I argued that the concept of accelerated backward time travel is perfectly logically coherent. We can, with perfect consistency, describe an adult who travels backward in time, let us say from the twentieth century to the seventeenth, and there lives several years, perhaps even the rest of his life. But traveling backward in time from the twentieth to the seventeenth century was not what Tim had wanted. Tim wanted to have been born in the seventeenth century and to have lived his entire life in the seventeenth century, having the experiences and the knowledge of a seventeenth-century man. He wanted to have had the memories of having grown up in the seventeenth century and to have had no knowledge whatever concerning what the future would hold for subsequent centuries. In short, he wanted *to have been*, not a time traveler to, but an inhabitant of, the seventeenth century. (The fact that he would not be alive in 1963 [the year in which we spoke of these matters], and indeed would have been dead for more than 250 years, did not trouble him in the least.)

Tim and I had not thought through his daydream in a careful, critical manner. We assumed that Tim was making sense, that what he wanted, although bizarre and physically impossible, was nevertheless logically possible. But his expressed desire was, even though the two of us may have thought otherwise, subtly incoherent on virtually any

viable account of what it is to be a person. (And that it stands as an object lesson in the possibility of engaging in incoherent discourse. Some self-contradictory desires are very unobvious: their incoherence emerges only upon thoughtful and deliberate probing.) And thus today, were someone to express a similar desire, I would challenge it on altogether different grounds: not on the practical or aesthetic grounds that in living in the seventeenth century one would have to forgo so much that is valuable in the twenty-first century, but on the logical grounds that no one alive today could possibly have lived in the seventeenth century. For Tim, or any other contemporary person, to want to have lived in the seventeenth century is of the same order of desire as wanting there to be a five-sided square or a colorless red apple. Such things, because their descriptions are self-contradictory, logically cannot possibly occur.

One way to focus on the incoherence is to try to imagine what it would be like, not for some anonymous other person to have lived his/her entire life in the seventeenth century, but for you yourself to have done so.<sup>14</sup>

Suppose this week an historian were to find both a portrait and a detailed diary of some seventeenth-century person. The painting is remarkable. It is of virtually photographic quality and it displays a person who, in outward appearance, is your physical double. And the diary is equally remarkable. It reveals a person who, knowing nothing what you know of the twenty-first century, reacts to the events and persons of the seventeenth century in much the way that you react to similar persons and events in the twenty-first.

Could this earlier person have been *you*? Suppose the current-you (i.e. the you alive today) had never been born. Would this earlier person, this seventeenth century look-alike and act-alike (to coin a word), have been *you*? Is it enough for a person to look like you and to act like you to really *be* (or to have *been*) you?

If you are not quite sure how you want to answer this question, try

---

14. I have often heard Professor Jonathan Bennett urge the ‘first-person’ test for various theories of personal identity. He cautions, for example, that one can imagine what it might be like for another person to undergo ‘splitting’ (mitosis), but one cannot, Bennett has argued, be so sanguine when it comes to imagining it of oneself. “Imagine your body undergoing mitosis during sleep. On which side of the bed would *you* wake up?” Bennett argues that *you* cannot imagine *yourself* waking up on *both* sides.

switching the centuries. Suppose, instead, that you were to be told by a seer, who has a perfect track record in all her short-term predictions, that sometime in the far future there will be a person who will look like you and who will act like you, but who will have no memories of you, or even for that matter any secondhand knowledge of your having lived earlier. Suppose you are inclined to believe the seer.<sup>15</sup> That still leaves open the question what you are to infer from her prediction. Would you regard this future person as *you*? Would you now feel that somehow you will escape earthly death to be reborn (resurrected, reincarnated, or what you will) in the future? Can you identify yourself with this future person, believing that that person is a future stage of *you*?<sup>16</sup> I think most of us will resist the suggestion that such a future person could ‘really’ be oneself.

If you do not share this intuition, then perhaps you might ask yourself how you would react to the news that someone who looks and acts just like you lives some four thousand miles away, right now, at this very moment. Suppose you were to meet that person. Would you be meeting a look-alike; or would you be meeting *yourself*? I think most of us, even if we were hesitant about the former cases – of the earlier and the later look-alikes – would be more reluctant still to acknowledge this contemporary person as being *oneself*. Our concept of *self* simply does not allow that we should learn that we exist not only here and now (e.g. in 2001 in British Columbia), but – surprise! – also at some distant place, e.g. in Moscow or in Paris. I may have a look-alike in Paris, he may even act remarkably like me, but whatever else is true of him, it surely is not that he is *me*.<sup>17</sup>

In saying this, I am of course appealing to your own sense of identity, and am assuming that, for most of us, our reactions and intuitions

15. If the spectacle of a seer is too much for your skeptical imagination, you might alter the example to that of a time traveler who has met this future person and brings back to you firsthand knowledge of your future look-alike.

16. If you are comfortable with the notion that time travel is coherent, then ask yourself what if this future person were to enter a time machine and were to travel back in time to the here-and-now and were to confront you face to face? Would you be shaking hands with yourself?

17. Recall Dickens’s *A Tale of Two Cities*. The case is more problematic, however, if the two persons share similar thoughts. Lorne Michaels (one of the producers of the television show “Saturday Night Live”) jokingly told the story that he had “become obsessed with the notion that somewhere in the

would be pretty much the same for the circumstances described. This is not to say that if your philosophical intuitions are radically different from mine then they are, somehow, wrong. The point of the exercise is not to judge a particular concept of personal identity right or wrong, but to try to bring into focus what one's concept of personal identity is. If your concept of personal identity is enough like mine to cause you to withhold identifying yourself with some former, later, or distant person just on the basis of similarity of features and of personality, then you, like me, will find Tim's request incoherent.<sup>18</sup>

For some person to *be* Tim (or to *be you*), it is not sufficient that that person share Tim's (or your) physical appearance and personality. Clearly something more is needed. (If not, then you could – even at this very moment – theoretically, if not in actual fact, exist at several widely separated places, e.g. London, Paris, or Moscow, having entirely different sets of experiences.) But I think most of us will be prepared to reply to the suggestion that we might be in several different bodies in several different places all at the same time by rejecting the suggestion, not as false, but as incoherent, i.e. as logically impossible. The suggestion is inconsistent with our concept of what it is to be a *person*.

### 12.7 What more might there be to the concept of *personal identity*?<sup>19</sup>

Tim's imagining that he could have lived in the seventeenth century overlooks certain ingredients which are essential to personal identity. That there might have been someone who looked and acted like Tim was certainly insufficient for that person to have *been* Tim. Something, perhaps a considerable amount, more is required for personal identity. I have earlier suggested that the 'something more' which is

---

world there was a person having the exact same thought he was at exactly the same moment. He decided to call that person, but the line was busy" ([95], 36, footnote). Would even this establish the identity of the two persons?

18. See (i) Thomas Nagel, "Death" ([141], 67) and (ii) Derek Parfit, "How Our Identity in Fact Depends on When We Were Conceived" ([149], §119, 351-6).

19. My thoughts about the topic of this section are in a state of flux and hence the discussion below is at best tentative. Thus, this particular section should be read with more than the usual degree of forbearance.

required includes veridical memory. Without veridical memory, there cannot be personal identity.

But are these two ‘dimensions’ all there is to personal identity? Does sharing veridical memories with, and having much the same personality as, an earlier person suffice to make the later person identical with the earlier? Many writers have assumed that veridical memory and shared personality comprise a set of sufficient conditions for identity of persons. But some writers are not satisfied with even these two fairly rigorous requirements; they believe that yet more, or something quite different, is required for personal identity.

Saul Kripke, for example, has argued that being born of the parents one actually has is a necessary condition for being the person one is. You could not possibly be identified with anyone having parents different from your own. Using Elizabeth II as his example, he writes:

How could a person originating from different parents, from a totally different sperm and egg, be *this very woman*? One can imagine, *given* this woman, that various things in her life could have changed [i.e. been different]: that she should have become a pauper; that her royal blood should have been unknown, and so on. One is given, let’s say, a previous history of the world up to a certain time, and from that time it diverges considerably from the actual course. This seems to be possible. And so it’s possible that even though she were born of these parents she never became queen. Even though she were born of these parents, like Mark Twain’s character [footnote: in *The Prince and the Pauper*] she was switched off [exchanged] with another girl. But what is harder to imagine is her being born of different parents. It seems to me that anything coming from a different origin would not be this object. ([116], 113)

For Kripke, Tim’s plaint – whatever else might have been incoherent about it – would have been impossible because it imagined that someone having different parents from Tim’s could, nonetheless, have been Tim. For Kripke, no one born in the seventeenth century could possibly have been Tim, since no one born in the seventeenth century was the child of Tim’s parents.

In a recent article in *Psychology Today*, Russell Belk, reporting on some recent experimental studies, writes:

What we possess is, in a very real way, part of ourselves. Our

thoughts and our bodies are normally the most central part of our self-concept. But next in importance are what we do – our occupations and skills – and what we have – our unique set of possessions. ... We generally include four types of possessions in our personal sense of self: body and body parts, objects, places and time periods, persons and pets. ... We found that academics were especially likely to cite books as favorite possessions, perhaps because they represent the knowledge on which their work is based. For other people, sporting goods represent what they can or could do ... Many studies have shown that the loss of possessions that follows natural disasters or that occurs when elderly people are put in institutions is often traumatic. What people feel in these circumstances is, quite literally, a loss of self. ([25], 51-2)

In Belk's view, for some persons the loss of material possessions (including external bodily parts) will constitute a radical discontinuity in *self*.

John Perry, in his estimable *A Dialogue on Personal Identity and Immortality* ([151]), presents to us the dying Gretchen Weirob. Her body has been fatally injured and will soon die. On her deathbed, she has been offered the choice of having her intact brain transplanted into the healthy body of a brain-dead patient. She refuses on the grounds that she cannot identify herself now with the future person who will have her brain but not her (present) body; that is, she has no *anticipation* of being that later person. Here Perry jolts our intuitions. Although he does not pursue the question explicitly, there is in the dialogue at least the suggestion that there is a certain *symmetry* between *anticipation* and *memory* in determining personal identity.

Virtually all discussions of personal identity involve cases of *re-identification*, i.e. of identifying later person(-stages) with earlier ones. But why this particular prejudice or bias? Why are there not equal numbers of discussions of *pre-identification*, i.e. of identifying earlier person(-stages) with later ones? Should the anticipating of being a future person – as sometimes occurs in discussions of eschatology\* – be factored into the equation of personal identity on an equal footing with memories of having been a past person?

As we collect these many suggestions – Kripke's, Belk's, Perry's, and others' – as to further (or different) necessary conditions for personal identity – having the parents one does, owning the things one does, having an anticipation of being some future person – difficult

and disturbing questions arise about the very practice itself of metaphysics.

There is always a desire in doing philosophy to construct economical theories, ones which seek the minimum set of conditions which are individually necessary and jointly (i.e. altogether) sufficient for the correct application of a concept. In analyzing the concept of *personal identity* it is natural to want an analysis which is as neat and tidy and free from loose ends as is possible. And thus there is a strong temptation to find grounds on which to reject most suggestions forthcoming as to further necessary conditions for personal identity. We want to be able to say: "These conditions, *x*, *y*, and *z*, are necessary; any further conditions are superfluous or redundant."

Thus, a while ago, when I and several of my colleagues were discussing Weirob's claim that she would not be identical with a future person who had her memories and personality but not her body, some of my colleagues argued that Weirob was simply mistaken: that when the surgery (brain-transplant) had been completed, and the patient woke out of anesthesia, that patient would recall having been Weirob and would insist that identity had been preserved. In short, these colleagues were prepared to tell the dying Weirob that she was *mistaken*, that she would survive if she would but consent to the surgery. In other words, some of my colleagues were prepared to place their own theory of personal identity above that of Weirob.

Is there some 'objective' theory of personal identity whose essentials might be grasped and the adoption of which would warrant our telling someone that he or she was wrong in conceiving of himself or herself in some particular way?

Not too many years ago, I myself argued in just the way my colleagues argued. I, too, believed that Weirob was simply mistaken: that hers was an *incorrect* view of personal identity, and that she had made a mistake not unlike believing that squares must be red or that material objects must be soluble in water. Her error, I thought, consisted in believing that some feature (anticipation, in this case) was necessary to the concept of *personal identity* when in fact it was not.

I no longer am so ready to insist on that particular view of the philosophical enterprise. Were someone to suggest that all squares must be red, I would be quick to object, arguing that that person had got the concept of *square* wrong. And the reason I would be comfortable arguing in that way would be because I do believe that the concept of square is fairly universally shared, that most of us do have virtually the identical concept of *squareness*. And in other cases, I might object

to someone's analysis – an analysis of *probability*, for example – on the basis that it was inferior to others or that it did not work particularly well, e.g. was confused or clumsy or applicable to too few circumstances. That is, in some cases I am prepared to argue that certain analyses of a given concept are better than, or preferable to, certain others.

But the concept of *personal identity* seems to me to be different. Indeed it now seems to me something of a mistake to talk or write about 'the' concept of identity. The more I read what other persons have written, and the more I talk with my students about their own concepts of personal identity, the greater looms the diversity between the many variants of the concept.

The concept of identity which is used in Law is probably a fairly minimal concept in that it invokes a minimal set of necessary conditions. (The Law could not function with a highly variable concept of personhood, no more than it could function with a highly variable concept of *property* or *responsibility*.) But this same concept may not be particularly useful, for example, in psychiatry, where a patient whose memory is intact may feel himself totally detached from earlier actions.

I think philosophers err if they believe that they can construct some one viable theory of personal identity. That particular goal is as illusory as trying to construct some one theory of, for example, what constitutes quality in music or beauty in art. The trouble is that if we look broadly across our culture, and particularly if we step outside it, we find immense differences in the prephilosophical intuitions persons have about personal identity.<sup>20</sup> The occupational danger for philosophers lies in our too often creating philosophy for other philosophers, indeed not even for all other philosophers, but only for philosophers who belong to the same 'school'. The hazards of inbreeding and of tunnel vision are ever-present.

I remain convinced that memory and personality are the essential

---

20. A great many articles and books published by psychiatrists and psychologists treating the concept of *self* arise out of their clinical experience with patients who have 'immature', 'defective', or even 'pathological' concepts of *self*. One must beware, however, not to draw from these writings the idea that the diversity of concepts of *self* arises out of arrested growth or psychological disorder. When we talk with persons whose concept of *self* is in no way dysfunctional, we find an equally prodigious range of difference.



*core* of the concept of personal identity. But I am no longer so sure that other factors might not also play an important role, and I am not confident that there are not, in fact, a great number of diverse, yet viable, concepts of personal identity, some of which are not merely different from one another, but even incompatible. In short, I am making a plea for tolerance in the matter of explicating the concept of identity. I think it hopeless, and indeed inappropriate, to argue that there is, or could be, one best concept of personal identity. Our concepts of personal identity are too varied to allow a single reconstruction.

With this said, I turn to our closing case study. If personal identity is not carried by a changeless soul, if personal identity requires continuity of memory and preservation of personality, and if personality, in turn, includes such things as intense desires and mental capacities, then there are some profound consequences in the changing of persons' desires and mental capacities.

## 12.8 Case study: Can there be justice after death?<sup>21</sup>

There are, I think, two principal egoistic motives which prompt us to desire an afterlife: a desire to maintain what is valued in our lives – including perhaps, but hardly limited to, the sensual, the intellectual, and the aesthetic – and a desire to achieve what we wanted but did not have in life – including perhaps, but hardly limited to, material goods, honor, power, creative talents, and physical abilities.

But for many persons, the desire that there be an afterlife is in part motivated by reasons which transcend individual, personal considerations. This world, we all know – and are constantly reminded throughout the day on the electronic news media and in the newspapers – is unfair. Indeed the world is grossly unfair. A catalogue of its injustices ranges widely from physical handicaps, sickness, and grief to starvation, slavery, flood, avalanche, wanton acts of terrorism, and so on so as to overwhelm the imagination.

Doubtless many of us find the notion of an afterlife appealing, not just because it holds out the promise of thwarting eternal personal Nothingness, but equally – and probably for some of us, even more strongly – because it offers the prospect of finally putting right the injustice in this world. It is in the afterlife, we have been so often

---

21. This section is a revised version of an essay which originally appeared in [104].

propagandized by religion, that virtue will be rewarded and evil punished. Our sense of morality craves this, whether or not it is in fact actual or even, for that matter, possible.

Usual philosophical discussions of justice concern the problems of realizing greater justice in this world. Such discussions typically rely heavily on specifying, and trying to work within, a variety of constraints: ignorance, scarcity of goods, legal systems, and – although often overlooked – physical possibility itself. Indeed so pervasive is the constraint of physical possibility, it hardly even is acknowledged. It is simply an unarticulated presupposition.

But what happens if one seeks to maximize justice in a world (e.g. the afterlife) which is not subject to these usual kinds of constraints? In an afterlife (heaven or hell or some other place), could the Dispenser of Justice (whether an individual or several minds working together) achieve perfect justice? What if, by simply willing it, the Dispenser of Justice could bring into existence any number and variety of goods? What if, that is, there were not scarcity but infinite plenitude? What if physical possibility were to become coextensive\* with logical possibility, i.e. the only constraint on the actual (afterlife actual, that is, not this-world actual) were the requirement that no self-inconsistency were to be realized? What if every veil of ignorance were to be lifted?<sup>22</sup> What if, that is, we should all know – if not everything – at least whatever we wanted to know? What if, in particular, every person's every deed were known? What if every person's every desire, doubt, hope, longing, envy, animosity, lust, love, were also known?

Could an omniscient, omnipotent Dispenser of Justice bring about perfect justice under these circumstances? Many religious believers, for millennia, have thought so. I find it difficult to share such optimism. Even in the afterlife, perfect justice – I am afraid – is unrealizable. My pessimism stems from several considerations.

---

22. Although I will not pursue the matter here, I must mention that an afterlife in which this world's physical laws do not hold true will present severe problems for epistemology. All human knowledge of contingent universal propositions presupposes a background of physical laws. Without there being a relatively fixed set of knowable physical laws, human empirical knowledge would seem to be significantly curtailed. In an afterlife where physical possibility expands to nearly the compass of logical possibility, a substantial part of human knowledge would have to flow from (what are in this world) unknown a priori sources, and not from a posteriori ones.

The principal difficulty, as I see it, is that persons sometimes have intensely painful desires which may be satisfied in but one single way. This is particularly awkward when the desire is not for a kind of material possession or a physical skill or a bodily appearance (a desire which the Dispenser of Justice could easily indulge), but for the company, love, or companionship of – not just some person or other, but – some particular person. How could the problem of, let us say, unrequited love be solved by the Dispenser of Justice in the afterlife? For it often happens that one person will form strong emotional bonds – of caring, of longing, of needing – to some particular other person, where the latter person loathes and actively avoids the former.

There seems to be no fully satisfactory, i.e. uncompromised, solution to this problem, although there are a number of apparent solutions. We might begin, for example, by arguing that justice does not demand the elimination of every possible pain. Justice, we might try to argue, demands only eliminating persons' pains when to do so does not infringe on the rights of other persons.

Are we then to ignore the pain of the person whose love is unrequited? Not much of a heaven, we might be inclined to protest, in which there is still so much pain. An innocent person whose love is unrequited might be suffering the pains of hell. How come this is permitted in heaven? What can be done to alleviate the undeserved suffering of this person?

The immediate temptation, since we are talking of heaven, where everything short of the logically impossible is possible, is to argue that the Dispenser of Justice could simply will away the sufferer's pain. If desire is causing intense pain to its owner, and if that desire cannot be satisfied because to do so would conflict with the rights of others, then it would seem that the next best alternative would be for that desire to be expunged, i.e. nullified, by an act of the Dispenser of Justice.

But the trouble with such solutions, and so many others which would have existence in the afterlife sanitized, sterilized, perfumed, rendered conflict-free, etc., is that they sometimes do violence to the very concept of personal identity. Consider the case of the parent whose entire reason for being is directed toward caring for and loving his/her daughter. But suppose the child reacts by asserting her autonomy. Above all else she wants to be free of, and distant from, her parent. Suppose, too, that these differences are irreconcilable. We might suppose that the parent's love in this case is overbearing; perhaps it is even irrational.

In a world unconstrained by physical laws, the Dispenser of Justice

could remove the parent's pain by eradicating the desire which engendered it. But would this be justice? We can imagine the prospect of having the painful longing removed being put to the parent before the Dispenser of Justice acted. And we can imagine the parent protesting: "To blot out this particular desire would be to destroy me. What makes me ME is my love and longing – however much grief it causes me – for my child. If I cannot have the love of my child reciprocated and you were to rob me of this pain, you will have annihilated me. This living body might remain, but whatever survives such a drastic alternation will not be ME."

Some persons do have such desires, desires which are intensely painful and yet which justice – because of the conflicting rights of others – cannot satisfy. But justice cannot always then fall back to a 'next best' solution, viz. eliminating the pain by nullifying those desires. For justice, surely, also demands the preservation of personal identity. And these latter two demands – the elimination of the pain of innocent persons and the preservation of personal identity – will sometimes be impossible to satisfy together. There are certain unsatisfiable desires, some of them intensely painful, whose elimination would be tantamount to extinguishing the person who had them.

There is a second sort of difficulty for the belief that justice might be realized in an afterlife. Do virtue and do wrongdoing have just deserts? Is the rehabilitation of the wrongdoer, is his/her contrition, is his/her restitution of wrongly appropriated property all that justice demands? Certainly for many cases the answer must be yes. But something deep inside many of us resists this answer for all cases. Some crimes are so heinous as to make rehabilitation and contrition wholly inadequate. Some crimes are of such magnitude as to make any thought of restitution insulting to the offended. Nothing humanly doable in this world could be fit justice for the crimes of, let us say, a Mengele, and he was – sad to say – not the worst.

For some crimes nothing short of punishment of the offender will satisfy the longing for justice the offended-against demand. But – and here's the rub – punishment often cannot be meted out to the guilty without causing pain to the innocent. I am not talking of the pain of persons opposed to punishment – although their pain may be real and deep – but of the more immediate pain of the wives, husbands, mothers, fathers, daughters, sons, friends, and lovers of the punished. Persons do not become guilty themselves and warrant punishment for loving a wrongdoer. And yet their pain may well be, very likely will be, intense when he is suffering his punishment.

Here, rationality takes a back seat to emotions. While knowledge that the child is in fact guilty and is being punished justly and deservedly may quell a parent's outrage and stifle his objections, that knowledge may do little to numb his anguish. Indeed, it may even make it worse by denying the parent the vent of a righteous fury.

Were the Dispenser of Justice to will away the grief and pain of the innocent person for his/her loved one's punishment, the grieving person would have been rendered less than human. Consider the case of a serial child molester and murderer. A very great deal of our horror, revulsion, and demand for his punishment is grounded in our empathy with the pain caused his victims' surviving families. That is, part of our outrage flows from our certain knowledge of the grief families will feel at the injury and death of one of their members. And yet when we demand punishment of the wrongdoer, his own – innocent – family will suffer because of his pain. What choices are then open to the Dispenser of Justice? Eschew punishing the guilty? render their innocent families insensible of the punishment? render their innocent families uncaring? The consequences of each of these alternatives seem to be forswearing justice, adopting subterfuge, and inducing callousness, respectively. None of these strikes me as compatible with perfect justice.

But this is hardly the end of the problem. There is yet a third difficulty. If there is an afterlife, what age are we each to be in that afterlife? Few, if any, nonagenarians would want to endure an eternity 'housed' in the body they had at the time of their death. No, justice would seem to require getting back your body when it was at its fittest and healthiest: a twenty-five-year-old body for most persons (that is, if the Metropolitan Life Insurance Company is to be believed in these matters). But what of the mentally and physically handicapped? What of the legions of children who died in childhood and never had an adult body? What bodies, what age, what capacities are these latter persons to have in the afterlife?

Surely the infant who died in this world at the age of two is not to remain an infant for eternity in the afterlife. Granted there are certain pleasures of childhood, but I think it the rare person who would willingly swap those of adulthood for those of childhood. But if the child who died at the age of two years in this world is not to remain an infant in the afterlife, what sort of person is he to be in that afterlife? Is he to mature in that afterlife, both bodily and mentally, as he would have done had he not died in this world?

One possibility would be to accelerate the two-year-olds immediately to adulthood. But this solution poses fresh problems of injustice. It strikes me as unfair to rob anyone of childhood. Having reached adulthood, I prefer it to childhood; even so, I would not want to have missed childhood.

If children who die in this world are to be given in the afterlife the childhoods they missed, the afterlife is going to have to resemble the planet Earth far more than it resembles traditional images of heaven. Put bluntly, heaven is no place for a human child to grow up. A place of plenitude and where physical possibility is – more or less – co-extensive with logical possibility just does not strike me as the fit playground for a young inquiring mind. On the contrary, it strikes me as a place where one cannot have much fun. To be a proper environment for a human child, the afterlife ought to be awfully like this world, complete with swings, trees, frogs (or similar sorts of *exotica*), schools, cuts and bruises, successes and failures, joy and heartbreak, etc. At the very least, it has to resemble Earth, not so much in appearance, but in physical law. It has to be a place in which much the same sorts of things have to occur as happen typically on Earth, and that means it has – of its very nature – to be a place where there is much unfairness. So while unfairness need not be a permanent feature of heaven, it must be of a part of heaven for at least as long as it takes the last-dead child to reach adulthood.

This still leaves the problem of the mentally handicapped. Are they in the afterlife to be made rational and intelligent? After all, it was unfair that they were not more rational and intelligent in the first instance. But how can rationality and intelligence be conferred on a severely mentally handicapped person without thereby destroying that person's identity? Marked increases in rationality and intelligence are certain to alter a person's personality radically: the desires, the expectations, the abilities, the typical reactions, the human relationships, etc. that the original person had are bound to change significantly. But these kinds of changes are just the sorts of ones which we regard as altering personal identity. A person who speaks fourteen languages, who runs a mile in 3:51 minutes, who discourses on the subtleties of Quine's philosophy, and who choreographs ballets to the music of Villa Lobos cannot in any but the most Quixotic sense be identified with an earlier person whose body he may have inherited, but who was deaf, dumb, halt, and incapable of understanding language.

My nagging fear is that the injustice which befalls some of us – par-

ticularly those so unfortunate as to be born profoundly mentally and physically handicapped – cannot be undone or recompensed.<sup>23</sup> The ‘not’ operative here is the ‘not’ of ‘not logically possible’, not ‘not humanly possible’. To undo a severe mental handicap is not to give someone something he lacked, but is to annihilate the one person and to substitute in his place another. Personal identity logically cannot be preserved over a change of this kind and this magnitude. The promise is often made by clerics that the injustices and suffering of this world will be ‘put right’ in the afterlife. But the promise is at best a falsehood or at worst a lie. There is no logically possible way to ‘put right’ the injustice of a person’s being born profoundly mentally handicapped.

In the end, I have a gnawing suspicion that the very existence of an afterlife is a myth. If it is not, then it is hard to see how it could even begin to live up to its billing. If there is an afterlife, it can hardly be the sort of place where justice is finally realized. The trouble is that justice logically cannot finally be realized. Not even a perfect Dispenser of Justice can bestow perfect justice on less than perfect beings, i.e. on the likes of you and me, our friends and loved ones, those we care about, and those we abominate. I find I am driven to agree with Boito’s Iago:

Man’s Fortune’s fool even from his earliest breath.  
 The germ of life is fashioned  
 To feed the worm of death.  
 Yea, after all this folly all must die.  
 And then? And then there’s nothing,  
 And heav’n an ancient lie.<sup>24</sup>

---

23. “... along with Helen Keller, my grandfather [Oklahoma senator Thomas Gore] was one of the most famous handicapped persons in America. We were very close. I was taught to read early so that I could read to him, and I read him the newspapers, the Congressional Record, history. When I was a little boy, a sob sister for a newspaper came to interview my grandfather. She said, ‘Senator, there must be so many compensations for your blindness, like a superb memory, sensitive hearing. Could you tell me what they are?’ And he said, ‘There are *no* compensations.’ That phrase has sounded continuously in my head ever since” (Gore Vidal; reported in [36], 53).

24. Arrigo Boito, libretto for Verdi’s *Otello*, act II, scene II, 1886