

An Interdisciplinary Journal of Philosophy

Editor: Alastair Hannay

Vol. 33, No. 1, March 1990

CARL PAUL ELLERMAN	New Notes from Underground	3
STEPHEN R. L. CLARK	Notes on the Underground	27
ARNOLD ZUBOFF	One Self: The Logic of Experience .	39
TORBJÖRN TÄNNSJÖ	Methodological Individualism ...	69
ANDREW MASON	On Explaining Political Disagreement: The Notion of an Essentially Contested Concept	81
WILL KYMLICKA	Two Theories of Justice (Brian Barry: <i>Theories of Justice</i> , Vol. 1 of <i>A Treatise on Social Justice</i>)	99
Books Received	121

One Self: The Logic of Experience

Arnold Zuboff
University College London

Imagine that you and a duplicate of yourself are lying unconscious, next to each other, about to undergo a complete step-by-step exchange of bits of your bodies. It certainly seems that at no stage in this exchange of bits will you have thereby switched places with your duplicate. Yet it also seems that the end-result, with all the bits exchanged, *will* be essentially that of the two of you having switched places. Where will you awaken? I claim that one and the same person possesses both bodies, occupies both places and will experience both awakenings, just as a person whose brain has been bisected must at once experience both of the unconnected fields of awareness, even though each of these will falsely appear to him as the entirety of his experience. I also claim that the more usual apparent boundaries of persons are as illusory as those in brain bisection; personal identity remains unchanged through any variation or multiplication of body or mind. In all conscious life there is only one person – I – whose existence depends merely on the presence of a quality that is inherent in all experience – its quality of being mine, the simple immediacy of it for whatever is having experience. One powerful argument for this is statistical: on the ordinary view of personhood it is an incredible coincidence for you (though not for others) that out of 200,000,000 sperm cells the very one required on each occasion for your future existence was first to the egg in each of the begettings of yourself and all your ancestors. The only view that does not make your existence incredible, and that is not therefore (from your perspective) an incredible view, is that any conscious being would necessarily have been you anyway. It is a consequence that self-interest should extend to all conscious organisms.

I. Against Tokens

I believe that we are all the same person. How could anybody come to think *that*? In this paper I shall try to explain how it happened to me.

In the late summer of 1961, just before my friend Bert and I entered high school, we were discussing a science-fiction story by Arthur C. Clarke called ‘The Other Tiger’. In the story, two fictional friends were talking about how, if the universe was infinite, there ought to be infinite duplications of every finite natural event down to every discriminable detail. For the number of discernible variations of such events, though incredibly large, would be only finite. Bert and I were then sitting at a small, round table in Sage-Allen’s Cafeteria in downtown Hartford, Connecticut. Not surprisingly, we began talking about the prospect of an infinite number of Berts and Arnolds sitting at small, round tables in an infinite number of Sage-Allen’s Cafeterias. I remarked that when we died there could still be

Berts and Arnolds carrying on. This thought had a certain sentimental appeal. But no more than sentimental, one might say, because such Berts and Arnolds would be mere duplicates and have as little to do with our lives as any other people who were so distant from us.

In early autumn of the same year Bert mentioned that he'd read an item about an attempt of Soviet scientists to exchange the brains of dogs. While I tried to picture to myself this odd undertaking, an even odder thought-experiment suddenly jumped into my mind. I imagined two brains lying at either end of an operating table. For the sake of vividness – please forgive me – let us say a mad scientist has only a moment ago snatched the brain from your head. It is one of the two on the operating table. The other brain is a precise duplicate of yours in every discriminable respect, including all its patterns of memory traces. Perhaps the scientist created this duplicate himself, or perhaps he stole it from the head of one of those duplicates of you that would have arisen naturally in an infinite universe.

Anyway, this mad scientist is capable of feeding into these brains any pattern of stimulation he likes, by means of electrodes plugged into them where nerves would normally be entering from the sense-organs and the rest of the body. And he has chosen to give both of them precisely the same pattern of stimulation that your brain would have been receiving if it had not been snatched from your head moments ago. That would be why it seems to you that your brain is still in your head, that my paper is still before you.

As I say, both brains are being fed exactly this same pattern of stimulation. What should we expect is true of the subjects and their experience? Would we not suppose that the episode of experience connected with each brain would be qualitatively identical? But would we not also think that, despite the completeness of their qualitative similarity, the subjects and their episodes of experience must be numerically distinct from one another? You are one subject, lost in one experience; at the other end of the operating table is another subject, lost in his or hers. It is as though we are thinking about two ashtrays of the same design sitting at either end of a coffee table.

But now for the experiment itself. Our mad researcher begins by trading one quarter of your brain for the corresponding quarter of the other. He does this instantly, or through instantaneous freezing and thawing, so that it does not register in the patterns of brain activity. What is the result? Surely it would not be natural to think that you have gone over to the other end of the table, along with a mere quarter of your material. Nor is it inviting to think that either you or your experience is a quarter changed in numerical identity. The natural thought is rather that you and the other subject have remained wholly what – and wholly where – you would have been if there had not been an exchange of quarters.

Next he exchanges a second quarter between your brain and the other's. Now, it was *wholly* you, in the same position, after the first quarter exchange, just as if it had never been done. So the second quarter must merely be like the first all over again. There is again no change in what or where you are. One thing to emphasize is that we may easily think of the brains themselves in terms of fractions. Thus, though we might be a bit puzzled about whether the *brain* that was originally yours is still the one you've got after the second quarter exchange, if we like we can just fall back on talking about there being half of the original brain with you and half now over there with the other. But one could never talk about the subject or his experience like that. So, anyway, what we should expect after the exchange of second quarters is that each subject is still wholly where it was at the start of the experiment.

You may guess what he does next. A third quarter exchange. And the result must again be the same; you remain where you are. And a final exchange, of fourth quarters, must also make no difference to where the subjects are. But notice that this final result looks as though it should be indistinguishable, except in history, from what would have been done if the mad scientist had merely picked up the brains at the start and exchanged their positions. All the material that was on the right is on the left and vice versa. But if this had been done all at once we would have been sure to say that the subjects as well as the brains had changed their positions.

The question is, could the difference in history that I have just indicated as the only difference between the two procedures be enough to make for the utterly different outcomes for numerical identity in the otherwise indistinguishable results? It seemed to me very implausible that the difference in history *could* be enough. What suddenly struck me was that the logic of experience was very different from what we normally supposed it was. In fact, when we were thinking about a particular experience we were, without realizing it, thinking about a type, a universal, rather than a token, a proper particular or individual. What I ended in thinking was that, in an experiment like ours, there had been numerically only one episode of experience and only one subject from the beginning. There was, despite the two brains at the two locations, but a single particular experience, as of being precisely you, in just that spot where you seem now to be, reading exactly like this, with all these current thoughts and sensations. And therein there was but a single you. All these specific qualities of the experience and its subject also determined the numerical identity. Experience and subject occurred equally well, numerically the same, in either of these realizations of them, in either brain. Therefore, at the end of either the series of quarter exchanges or the immediate whole exchange, there was no distinction of a subject on the right in contrast to a subject on the left. No matter which way one did the exchange, there was throughout but one subject – you – possessor of but a single experience.

Let us compare the logic of experience to the logic of something like a novel. A novel might be called a 'detailed type', of which there are 'tokens', which are its copies. For example, on a shelf in a bookshop there might be two copies of but a single novel, *The Adventures of Huckleberry Finn*. Just as this would be only one novel, this would also constitute no multiplication of the character called Huckleberry Finn, despite there being two copies of his adventures on the shelf. The logic of a copy is different from that of a novel. If one of these copies was destroyed, the novel would continue to exist in the shop so long as there was at least one copy there. The novel has the logic of an Aristotelian universal. There must be at least one instance for it to exist, but repeated instances cannot multiply the number of universals.

One could carry out on this novel essentially the same operation that I described in my thought-experiment. After buying the two copies, one could cut each of them down the spine between the same pages at the middle of the novel; and one could then exchange the first halves of each copy. What would have happened? It would have affected the identity of the copies. Half of the copy that was on the left would now be on the right. But the identity of the novel would not have been affected at all. The same would have been true of the character Huckleberry Finn within the novel. His identity would have been unaffected. By the way, I have actually performed such actions during talks I've given on personal identity; but, since in a thought-experiment the experiment is in the thinking rather than in the doing, my audiences and I can claim no advantage in deciding these questions over people who have never seen this wonder but only imagined it.

Consider a land whose inhabitants have never known there to be more than one copy of any book. This limitation might result in a rather stunted appreciation of the logic of a novel as distinguished from its copy. To make these people aware of this distinction, a philosopher might invite them to participate in some thought-experiments. He would ask them to think what they should say if someone were to duplicate one of their books. Couldn't they then distinguish the novel, which would be numerically the same in that new copy as it was in this, just as already they had distinguished a word from its instances? Then it would turn out further that the novel could continue to exist when this mere copy was destroyed, so long as any other copy still existed, that the novel could continue to number one when there were two or more copies of it, that whether or not a book was exciting would depend on the novel rather than the copy, and so on.

What I am doing with regard to an experience is something similar to this. I am raising the possibility that what we think of as a particular experience is a concrete universal, like a particular novel, and remains numerically one if instantiated more than once, as in the case of the

duplicate brain activities in my story. I am raising the further possibility that the subject of the experience remains numerically one in both such instances, just as Huckleberry Finn, the character, is numerically one no matter how many copies of his adventures are produced. Like that of a novel and its character, perhaps the identity of an experience and its subject depends merely on a pattern and is therefore indifferent to changes in the particularity of the medium in which that pattern is maintained. And perhaps the particularity felt in an experience, the *this* and *here* and *mine* in it, is merely the subjective impression existing equally well in every occurrence of that pattern, like the setting and perspective of a novel.

Recall that the idea of the brain exchanges came to me in the early autumn of 1961. Soon after that I became anxious about death. One day in my high school Latin class the teacher told us about how once he had watched someone dying calmly by the side of a highway. At home that afternoon, I saw a war film on television. Soldiers were going, one-by-one, against a machine-gun nest. I was thinking what my own life was for me – a whole rich world in its own right. But each time a soldier fell to that machine gun's fire such a life turned to nothing – eternally. What bothered me especially was trying to understand this eternal non-existence that each of us faced. What would it be like to be dead? It couldn't be like the experience of darkness or blankness, because there would be *no* experience. And I couldn't look ahead to my death as I would a temporary episode of unconsciousness, since there'd be no experience of waking to follow it. This was my own eternal future; yet of it I could make nothing.

Perhaps under the pressure of this concern, the two reflections I opened by describing, on duplications across the universe and on brain exchanges, suddenly merged in my mind to give me a hope of life continuing, an argument for immortality. The Berts and Arnolds in Sage-Allen's Cafeterias, scattered through the universe, should it be lavish enough to repeat them, would not be mere duplicates of each other. An experience of being Bert or Arnold would exist numerically the same, regardless of where or in what particular organism it occurred, much as Huckleberry Finn in his adventures would exist in any copy of the novel anywhere in the world. If one copy of *Huckleberry Finn* is destroyed, Huck Finn continues to exist in copies that are still around. And I, a subject of experience, would not suffer the fate of a single copy. I would endure like the novel and its character.

Think of just one place in the universe in which this, your experience of reading, is instantiated. Now imagine that all life in this place is about to be extinguished by the sudden impact of a giant meteorite, in the middle of your reading of the next sentence. Well, if anywhere in the universe there is this experience of being *you*, *here*, and reading further, then you and your reading simply *do* continue, just as now. Your experience could

thus be continued regardless of *any* such local extinctions, in a generous enough universe.

This was very exciting. I had discovered a chance for immortality, or at least for an indefinite extension of life.

But what about the objective timing of such an extension? Surely this could make no subjective difference. Your continued reading could take place billions of years before or after the instantiation destroyed by the meteorite and yet it would be experienced simply as this same continued reading. An experience must be a universal across times as well as across brains. This experience of being you, here, *now*, would be numerically the same *whenever*, as well as *wherever*, it was realized.

So, according to my thoughts in 1961, the particularity of tokens, their particular places, times, and parcels of matter, was irrelevant to the particularity of the subject or of his experience. On the ordinary view, when identical twins are created by the splitting of an embryo, they will be distinct persons even though their biological type is the same. The crucial difference between them, then, seems to be one of token, of the particular material that forms them. Of course, they will also then come to differ from each other in their detailed psychological and physical descriptions. But on the ordinary view, surely, even if these details of type could be controlled so as to be kept precisely the same, as they were in our thought-experiment of the precisely duplicated brain or in the possible duplications across the universe, the difference in token would still be enough to distinguish two persons. Yet, on this same ordinary view, the mere existence of the particular material of such a token would not have been enough in itself to make for the existence of the person in question. That person would not have existed unless the right type, biological and perhaps psychological, had also been impressed on that matter. If, for example, the particular matter on which we would base the distinction of you from a twin had simply remained scattered in uneaten food so that your mother had never obtained it for use in the formation of her child, *you* would never have seen the light of day.

The usual view, then, requires that both a specific type and a specific token meet in a type–token match if any given individual is to come into existence. There is a tight specification of both type and token, in an a priori formula, dictating the necessary conditions for the existence of each potential individual. But this tight prefiguration of a person's type and token in such a package of necessary conditions leads to two sorts of problem for the ordinary view. The first sort of problem is conceptual, and we have already caught a glimpse of it in the inadequacy of the usual view in dealing with questions of identity in the thought-experiment of brain exchanges.

The second sort of problem with the ordinary view is that it makes it so

improbable from your own perspective that you should exist. For anyone else you would be just another type–token match among those that nature happened to produce. But for you this particular type–token match is special in being the only one that could have brought *you* into the world. The fact that this match was special for you, independent of whether it was brought into reality, makes it an enormously lucky outcome for you that nature included it among the incredibly small fraction of potential matches that *were* realized. How much more likely it would have been that the particular atoms necessary for you had been scattered somewhere else in the universe, rather than collecting just where they were needed if *you* were to be formed instead of one of the countless potential twin-like people who could have come about in your place with a difference in token. On my view, of course, any of these would have been you. No luck was needed. The result that you find, that you exist, would thus be no surprise.

Imagine you walk into a room and find a thousand coins lying on the floor, all of them heads-up. You are not allowed to examine these coins, but later you meet two people who give you different accounts of what happened. One says he knows that your host dropped a box that held all double-headed coins, which would have made the result you saw inevitable. The other says he knows that the dropped box held all fair coins, which would have made this result extremely unlikely, with a chance of only 1 in 2^{1000} . Because the fair-coin theory includes the fair-coin improbability of all thousand coins landing heads it is, analytically, to that extent improbable. You must infer that it is 2^{1000} times more likely that these are double-headed coins.

It was twenty years after my first thoughts about the conquest of death that I was struck by the idea for a statistical argument similar to the one I've just described, an argument of extraordinary power that could establish the irrelevance of tokens to personal identity beyond any reasonable doubt. The heart of this argument was already present in my thoughts about death.

For in these thoughts it is that your identity does not depend on the fate of a particular token that makes it easier for you to survive, to remain in the world. And it is this same accidental relation to the token in which you were realized that also made it overwhelmingly more likely that you had come into existence in the first place. You know that you *do* exist, and therefore you should infer the overwhelmingly greater likelihood of the theory that made this known fact itself overwhelmingly more likely. You can thus use the fact of your own existence as tremendous evidence against the relevance of tokens to your identity.

Imagine that you are awakened from a deep sleep, and your mother explains that you must fetch a copy of *Huckleberry Finn* from the local bookshop. It seems that one of two visiting aunts has phoned the shop and discovered that they have a copy; otherwise you would not have been

awakened. But your mother hasn't mentioned which aunt phoned the shop. Bella is aggressive and direct. She would have demanded that the shop assistant find the novel no matter where it was in the store. Tilly is incredibly shy. You know for sure she would not have wanted to hear about whether the shop contained *Huckleberry Finn* unless it was the one book copy nearest the phone. If Bella had been responsible, your awakening would have depended on the existence in the shop of a type, the novel. If Tilly had phoned, your awakening would have depended on the improbable match of the type, *Huckleberry Finn*, with an independently specified token – the book copy nearest the phone. You *are* awake. Hence you may infer the far greater likelihood that Bella was responsible. that your awakening depended on a type, not a token.

II. Against Types

At the beginning of this paper I said that I believed we were all the same person. So far I have described how I came to believe that there would be but one person across a multiplicity of tokens. I have, however, kept aside the question of whether different *types* would distinguish different persons. In both the brain exchange thought-experiment and the consideration of duplicate organisms across the universe we had stipulated that the type would be the same, not only the biological and general psychological type, but every detail of the content of experience. Was such a strict sameness of type really necessary for a sameness of person, even though sameness of token was not?

Let us focus on the question of whether a specification of the content of an experience is necessary to the identity of its subject. Recall the duplicate brains of the brain exchange thought-experiment. Imagine again that they have both been involved in the same detailed experience, as of you now reading this paper. But then, let us say, one brain is made to register the sensation of a jab in the back, while the other brain registers the welcome tones of a familiar voice. We have seen that the distinct brains, the tokens, do not make for distinct experiences or subjects. But does not a difference in experiential content such as the one I have just introduced bring about such a distinction? For how can one subject and one experience remain the same with two mutually excluding experiential contents?

The problem raised here is the great puzzle that confronts us when we consider actual cases of brain bisection.

If a surgeon cuts the direct connection between the hemispheres of the brain, at the corpus callosum, epileptic seizures are prevented from spreading from one side of the brain to the other. Such brain bisection seems not to interfere with a patient's normal functioning, since the halves of the brain are usually receiving similar information from the senses and

may continue to work, therefore, in relative harmony. But in a series of experiments in the 1960s, with split-brain patients as the subjects, the experimenters kept separate the information let into one of the hemispheres of the subject from that let into the other. The challenging result was that the subject would seem then to possess two mutually excluding experiences at one time.

Gazzaniga and LeDoux, in their book *The Integrated Mind*, describe this result as follows:

One of the immediate and compelling consequences of brain bisection was that the interhemispheric exchange of information was totally disrupted, so that visual, tactual, proprioceptive, auditory, and olfactory information presented to one hemisphere could be processed and dealt with in that half-brain, but these activities would go on outside the realm of awareness of the other half-cerebrum. Thus, the data confirmed the earlier animal work by Myers and Sperry but were, in a sense, more dramatic, in that only processes ongoing in the left hemisphere could be verbally described by the patients, since it is the left hemisphere that normally possesses the natural language and speech mechanisms. Thus, for example, if a word (such as *spoon*) was flashed in the left visual field, which is exclusively projected to the right hemisphere in man, the subject, when asked, would say 'I did not see anything', but then subsequently would be able, with the left hand, to retrieve the correct object from a series of objects placed out of view. Furthermore, if the experimenter asked, 'What do you have in your hand?' the subject would typically say, 'I don't know'. Here again, the talking hemisphere did not know. It did not see the picture, nor did it have access to the stereognostic (touch) information from the left hand, which is also exclusively projected to the right hemisphere. Yet, clearly, the right half-brain knew the answer, because it reacted appropriately to the correct stimulus.¹

Let's explore the paradox in this result for the usual view of experience.

There are now two separate streams of conscious experience where before there was one, this one having belonged quite straightforwardly to just one person, as your experience now belongs to you. We might consider that the division of the stream occurred immediately following the operation, even when relative harmony in hemispheric activities was maintained through shared stimulation. Or we might consider the separation of experiences to have occurred only during the experiments, when the activity of each hemisphere became so alienated from that of the other. Either way, we are faced with the question of the relation of the original person to the two resulting streams of experience. Can we think of these as both at once belonging to the same subject of experience? But each experience excludes the other. It seems there must then be an actual doubling of the subject, each of the two having its own experience. But if there are now two subjects of experience rather than one, each functioning through a different hemisphere of the brain, how are they related to the original subject that functioned through the undivided brain? Are they somehow both the same as the original subject? But they are not the same as each other. The

original cannot then have continued the same as both. Is just one of them the original? Then, which one? Consider how a stroke will sometimes leave a person with only one functioning hemisphere. Whichever hemisphere remains, we regard the person as living on through that hemisphere. The nature of his impairment, of course, depends on which hemisphere he has lost. But *he* is there to experience the impairment either way. In the split-brain case there are still both hemispheres, either of which would have been sufficient to carry on for the original subject of experience if the other had been lost. Having both is an embarrassment. As we have said, it seems the subject cannot continue in both. But there is no way the subject can be seen as existing in one of the hemispheres and not the other. Finally, it seems equally wrong to say the original subject has disappeared and these are two fresh subjects, since, once again, either of those hemispheres would have been sufficient for continuing his life; and here we have both.

It seems we cannot say anything about who these subjects are, and yet it seems there must be an answer to the question. Either the original person continues or not, and it is important to him whether and where he does.

This is the same paradox as that evoked by the notorious philosophical puzzle cases of human fission or fusion. In these thought-experiments people divide like amoebas, the hemispheres of one person's brain are transplanted into different bodies or the products of such divisions are united once again. The unanswerable questions are the same as those in brain bisection. Let me mention another instance of this puzzle as well, one which is neither fanciful like the usual puzzle cases nor rare like brain bisection.

Consider how it could have happened, though it didn't, that your embryo had split and two or more people, twins, had then developed from the products. Would you now have been all of these people, one of them or none of them? Notice, by the way, how such fission cases can be turned into fusion cases as well, just by reversing one's perspective. Thus here we may ask whether any of those persons who would have been produced by the splitting of your embryo are alive and conscious in you now. All of them? One of them? None of them? All answers seem wrong. But there must be some answer.

Let me bring out this paradox more explicitly with regard to the integration of the contents of experience by using a variation on a thought experiment in Derek Parfit's paper 'Personal Identity'.² Imagine that your own brain were so fixed that by pushing a button you could anaesthetize your corpus callosum, rendering it temporarily incapable of carrying impulses from one hemisphere to the other. On the radio tonight there will be a wonderful concert you've long wanted to hear. Yet you must at the same time be studying some dreary material for a course you are taking. You push the button after arranging that the concert will go into only the

right and the studying into only the left of the temporarily separated hemispheres. My question is, what kind of time will you have? Will you both enjoy the concert and suffer through the studying even though each of these experiences excludes the other? Will you only enjoy the concert or alternatively only suffer through the studying? Which one? Will you have no experience? But after a stroke, you would certainly have had whichever experience was in the remaining functioning hemisphere. In our story there is too much of what you need, not too little. There must be an answer, but it seems there is none.

I think the answer is that one subject *can*, in a single next moment, experience two differing non-integrated contents of experience. *You* will experience as *next* both the concert and the studying. Look once more at the latest version of the brain exchange thought-experiment. Recall that at first both brains were involved in precisely the same pattern of experience, as of you reading this paper. We established before that this was numerically but one experience and subject, despite the numerical distinctness of the brains. But we then further imagined a next moment in which the content associated with one brain featured the sensation of a jab in the back, while that associated with the other featured instead a pleasing voice. Which do you, the original single subject in both brains, experience as your next moment? We know that the identity of the brain can make no difference to this. The experiential effect must be the same whichever brain entertained either content. The objective time of the experience is also irrelevant. Either content would simply be experienced as having been *next*, as *now*, even if it were to be instantiated a billion years later or earlier than the other. So what *is* the experiential effect?

Both the experience including the jab and that including the voice involve the same abstract properties of being '*mine now*', despite the difference in content. And what thus makes any experience seem to be both *mine* and *now* is really but a single abstract quality, one possessed by all experience, anywhere, anytime. That is the quality of the immediacy of the experience – the immediacy of the experience for whatever is having it, in whatever time it is had. The experience including either the jab or the voice must have this same quality of being *mine* and of having been *next* (in other words, of being *now*). And this is all it will take for both patterns of experience genuinely to be *mine now*. Experiential content, the precise pattern of experience, is a *detailed type*, as we learned in the first talk. The identity of the experience itself, however, of that which is *mine now*, is dependent on only the *abstract* quality of immediacy present in all the detailed types of experiential content.

All experience, in all conscious things, has this same immediacy. All experience is *mine* and *now*.

In a relevant passage from Thomas Nagel's paper, 'Physicalism', he asks us to

consider everything that can be said about the world without employing any token-reflexive expressions. This will include the description of all its physical contents and their states, activities and attributes. It will also include a description of all the persons in the world and their histories, memories, thoughts, sensations, perceptions, intentions, and so forth. I can thus describe without token-reflexives the entire world and everything that is happening in it – and this will include a description of Thomas Nagel and what he is thinking and feeling. But there seems to remain one thing which I cannot say in this fashion – namely, which of the various persons in the world *I* am. Even when everything that can be said in the specified manner has been said, and the world has in a sense been completely described, there seems to remain one fact which has not been expressed, and that is the fact that *I* am Thomas Nagel. This is not, of course, the fact ordinarily conveyed by those words, when they are used to inform someone else who the *speaker* is – for that could easily be expressed otherwise. It is rather the fact that *I* am the subject of *these* experiences; this body is my body; the subject or center of my world is this person, Thomas Nagel.³

And if one's description includes the world at all times, there would seem also to be the additional fact that only one of these times was *now*. My view is simply that it is an illusion to suppose there are such additional facts. There is no exclusive *now* centring on one time and no exclusive *mine* centring on one organism. What I advocate is a demystification of our ideas about personhood and experience. All the occasions of consciousness feel as though they are exclusively *mine now*. None of them really is exclusively so. There is but one quality of being *mine now* existing in them all. And the quality of *mine* and *now*, as I have already remarked, is really just that quality of immediacy which is present in all experience. Thus personal identity depends on an abstract rather than a detailed type; it depends on this abstract universal of immediacy. I'm not saying that you *are* this abstract type. I am saying that any being that is capable of experience with immediacy – and hence any being that is capable of any experience at all – has everything it takes to be you. This view might be called 'universalism'.

But, one may object, how can all experience belong to me without my knowing it, without an integration of all mental contents? It may be helpful, I could respond, if we consider a form of aphasia that occurs when a lesion in the brain interferes with the connection between the visual cortex and the speech centre.⁴ The patient will say he is blind or make other inappropriate reports of what he is seeing. For there is no way his speech can be integrated with his visual processing. Yet he will otherwise function as though there is nothing wrong with his visual awareness. He will move himself about and manipulate objects with the full aid of his eyes, since the connection between movement and his visual processing is unimpaired. It is difficult to think of it being one subject with this strange disruption of the normal integration of mental functions, at once both seeing and not seeing. But how many subjects are there? With brain bisection, whose puzzle is essentially the same, there is at least an initial temptation to say

there are two subjects. Here even that is not tempting, because in these cases of aphasia speech is still connected to the movements responsive to vision. This puzzle is solved only when we realize that, along with giving up the mistaken individuation of experiences and their subjects by tokens, we must give up a mistaken individuation of these according to the connections of experiential contents. This is false even to the experience of normal mental functioning, where it is impossible that all content be thoroughly integrated. My sensing of my feet and my writing these sentences – how thoroughly integrated are they?

Even on the ordinary view of an experience, one may distinguish its particularity from its content. My experience now might have had a different content if things had gone differently. If I had turned my head upwards, I would now be seeing a ceiling instead of this paper on which I am writing. My claim is that one experience, one state of consciousness, is capable of having any sort of content without any requirement of integration. Psychological integration must be rejected along with sameness of token – in favour of the abstract property of immediacy – as the criterion of identity for both experience and its subject. Numerically the same *I* and *now*, the same *mine* and *this*, exist in all unjoined contents, in all separate tokens, including what we are used to thinking of as exclusively your life and body and exclusively mine. Thus all the lives and bodies of conscious beings are equally yours and mine.

Let's return for a moment to our literary analogy. Experience and its subject depend not on a concrete type, like *The Adventures of Huckleberry Finn*, but on an abstract type, like *writing* – which remains numerically the same in all the varied productions of writers and in all the copies of these productions. When I claim that, throughout the various bodies, minds, and experiential contents of all the world, there is but one self and one consciousness, I am not positing the existence of a strange new substance, any more than I would be positing the existence of a strange new book if I claimed that various books were all instances of writing. Neither am I putting forward a thesis with what might be called a crude causal significance. I am certainly not claiming, for example, that experiencing beings can somehow communicate in an extra-sensory way just because there is this same *I* and *now* in all their experience. Thus, although I am saying that after brain bisection there is only one next experience belonging to one subject in both disconnected hemispheres, I am not saying that this metaphysical singleness will somehow mechanically integrate the neural activities or the experiential contents. These will remain just as separated as they would have been when regarded as involving two subjects with distinct experiences. There are large consequences flowing from universalism, but none are of this crude mechanical sort.

Why might we tend to think of the integration of mental activity as

essential to the identity of a person? Imagine a world in which people have only ever known one object with the colour red. Such naïve observers might easily fall into thinking that to be red somehow also required having the other particular features of that one red object.

Similarly, the universal immediacy of consciousness is only ever experienced within the various limits of particular mental integrations. The one self, though defined by the quality of immediacy alone, always finds itself seemingly bounded by limits of mental activity, by limits of current and remembered experiential contents that happen to come packaged together. Within the reach of one nervous system and the memory it supports lies, seemingly, all the experience that has or has had the intimate quality of being *mine*. But accepting this suggestion is as clumsy a mistake as thinking that there can be only one red object if only one has been seen.

After all, the reach of integration cannot actually *make* experience mine. Integration can only *reveal* experience as mine. Recall your experience of the previous sentence. It was yours; it had the required immediacy. But obviously there is also unremembered experience, which is therefore unintegrated with the present content of experience but which nevertheless had the same quality of being yours. Think of the experience of the fifth bite of breakfast 38 days ago. It has little integration with the present content of experience; it has dropped out of memory; but surely it still had all the immediacy that would have made it yours. Much more startlingly, the same is true across the nervous systems of different organisms or, as in cases of aphasia or brain bisection, across non-integrated parts of a single nervous system.

What at last convinced me of the irrelevance of any detailed types to personal identity was the discovery, as late as January, 1983, of a statistical argument that opened up types as convincingly as the earlier argument had opened up tokens.

Suppose for a moment that your existence had required a detailed type, such as a particular pattern of experience, memory or genetic coding. Then there would have been an enormous coincidence attached to yours having been a pattern that occurs naturally. Of all the types that might have a priori defined someone's identity, yours would have happened to be one of the incredibly small proportion reflective of the actual order of nature. Contrary to the popular saying, it doesn't take all kinds to make a world. Most logically possible psychological or biological types don't come anywhere near being among those that make this world. Think of a person who would only come into existence with a set of memories that reflected years of flying about on winged horses. Well, this requirement would have been far closer to the actual way of things, and far less chaotic, than the requirements of the overwhelming bulk of potential subjects of experience, as these might have been individuated by logically possible patterns of

memory. Yet, we are asked to believe, the pattern of memory required by *your* necessary formula just happens to have fallen within the incredibly narrow bounds of memory patterns reflective of the actual world.

Compare this with a view that abandons such detailed type prefiguration. For such a view the fact that the content of your experience reflects the natural world is itself explained quite naturally. This that you experience is the sort of content the world itself provides; and you, free of any a priori constraints, will simply receive any flavour the world might be serving. There is no coincidence, then, between your a priori type and what the world will supply. Your type is abstract enough to allow for whatever comes along. Thus the constraint that makes experience reflective of a natural world such as ours is located where it belongs – solely in the character of the world itself. We must reject the incredible coincidence of the first view in favour of such an explanation.

Similarly, we must junk any view that imposed a biological type as a condition for your existence, a type such as the pattern of your genetic material. Think of all the logically possible biological types; an overwhelming proportion of them would never develop in the actual world. If we accepted biological type prefiguration, it would be outrageous luck that you were defined by a proper pattern of genetic material.

My alternative, once more, is to see your existence as indifferent to all such details. The abstract type of immediacy of experience is all that's required for your presence in the world. You will, therefore, simply find yourself with whatever biological types nature employs in producing its conscious beings.

Imagine you have been given the successful password for entering a room. The only two available accounts, let us say, are that this password had been assigned to you arbitrarily from among the typings of a monkey or that it had been obtained from the doorkeeper himself. Given its success – and given that it is a proper word – which is more likely?

The overall conclusion concerning both tokens and types is that any token or type that is supportive of experience is consistent with your existence. There is therefore nothing remarkable about the fact that you exist. Just as the usual view takes it as accidental to your existence what kind of clothes you are wearing, my view characterizes as also accidental all the other specific conditions of your existence. There is otherwise no disagreement between the views about the conditions that do obtain. The one condition of immediacy of experience is all that is necessary, on my view, for the haver of experience to be at once you, I, and everyone; and all the specific conditions defining all the individual organisms in which we find ourselves are irrelevant to our existence, and are for us no lucky coincidence.

III. The General Statistical Argument

Let's think about luck. Consider a particular game of roulette. Imagine that the only player at the table puts his money down only once and on only one number and then that this number happens to be the very one at which the little ball stops. That was both a coincidence and, since something good depended on it, luck. It would have been neither coincidence nor luck if the ball had somehow been made to roll to that number because the player had bet on it or if the player had been allowed to place his bet after seeing which was the winning number. Why would such cheating not have been luck? The relevant property of a number, the ball's landing on it, belongs to only 1 in 37 numbers. A specification of only 1 of these numbers by means of a property that is *random* relative to this winning property, such as that the number was the first to pop into the player's head, has only 1 chance in 37 of naming the number at which the ball actually stops. In a fair win, such a random designation coincides with possession of the winning property by mere chance. The slimmer the chance, the greater the coincidence and the luck. If, however, the designation of the number was not really random relative to the ball's landing on it, as in cheating, there are no odds of 1 in 37, no coincidence and no luck. It would be fallacious to argue that since the number on which a cheater bets still had only a 1 in 37 chance of coming up, it is still a coincidence that it does so. Coincidence requires that there be two or more independent designations of one among many items. The bet that is fixed by the outcome of the spin or the outcome that is fixed by the bet is really just a single specification of a number. The odds of such a match were 1 in 1. When we say a roulette number has only 1 chance in 37 of coming up, we must mean the number as designated randomly relative to whether or not it will win. Similarly, it would be fallacious to argue that there is anything coincidental in a number coming up when it has not been bet on, called or otherwise particularly designated in a random fashion. Without some sort of specification that is random relative to its coming up, it is merely the number that does come up. As such, it had the probability 1 of coming up. To think otherwise could be called 'the fallacy of the false coincidence'.

It seems a coincidence in anyone's eyes when the only player, betting only once, has picked the winning number. But imagine there are players betting on most of the numbers. When one of these players wins, that is no longer a coincidence for the uninvolved observer. Solitariness cannot underline this winning bet in a way random to its winning. But for the winning player his betting on the winning number still *was* both a coincidence and luck. We might be tempted to call this a 'relative' coincidence. For this winner *is* singled out *for himself* in a way random to his winning, in that he must be but this one same special player, himself, win or lose. (Notice that for the purposes of this discussion we are working with an

assumption of the truth of the ordinary view of personal identity.)

Imagine that a different film was made of each player and you were shown only one of these films at random. If that happened to be the film of a winning player, then that would be a coincidence for you. This winning player, and through him the winning number, would have been randomly designated for you by the random selection of the film. If, however, you are to be shown a film of a winning player which has been selected for your viewing *as such*, there can be no coincidence for you in seeing either the winning player or a winning number. We may call this a case of a 'falsified' coincidence – falsified for you, but still just as true for the winning player, because it was still selected for *his* observation randomly relative to the winning. This winner knew the event under the randomizing condition of its simply being *his* bet, win or lose. Your observation here, however, involved no such randomizing condition. By the way, what may have seemed to us an 'absolute' coincidence – a solitary player betting once and winning – could also be falsified for you if you were shown a film of such an event which had been selected *because* it showed a solitary player betting once and winning. So even such a coincidence was really a relative one; it needed an observer who saw it under a randomizing condition.

Now let's turn to your luck in existing. Consider your begetting, when a sperm cell met an egg cell to form the cell called a 'zygote' that developed into you. A conservative figure for the number of sperms competing to reach the egg in a conception is 200,000,000. On the ordinary view, not you but a potential brother or sister would have come into being if any other of those sperms had reached the egg before the one that did. The odds against the creation of any one of the potential individuals represented by all the possible combinations of sperm and egg are, then, 200,000,000 to 1, roughly the same as the odds against winning the single grand prize in a national lottery entered on equal terms by most of the adults in the United States. Now, if I am introduced to the winner of such a national lottery merely under the condition that he *was* the winner, there is no coincidence there for me. The absence of a randomizing condition in my observation has falsified for me any coincidence involved in the winning. If I had met a lottery contestant under a randomizing condition, however, such as that he was the nearest to me when I first inquired about who had won, it *would* have been a coincidence for me that this turned out also to be the winner. And his winning must similarly be a coincidence for him, since he discovers himself always under a randomizing condition relative to winning a lottery, that he is truly the nearest to himself – he *is* himself. He would be himself whether he had won or lost the lottery, so in his also having the winning ticket there is, for him, a genuine coincidence – and genuine luck as well. Now, I cannot meet you, winner of your conception, under any such randomizing condition. Any person I see as the result of that begetting *is* simply he or she who has won that strange lottery. If I

could somehow have designated at random a particular sperm before the conception and it had then turned out that the person that actually resulted from the conception was the one individual dependent on that very sperm, then there would be in this a coincidence for me. And if for some reason *my* life too depended on that outcome, here there would also be *luck* for me. But, again, I only ever see the winner *as* such. But *you were* dependent on only one possible combination among the potential 200,000,000 in your begetting. You weren't merely going to see the winner no matter who it was (as *would* be the case on my view of personal identity). Your dependence on just one of these possibilities is the randomizing condition that makes the resulting combination an extraordinary coincidence for you – and enormous luck too, if you value your existence. Therefore we may call the coincidence for anyone, on the usual view of personal identity, that his was the winning combination in his begetting an 'egocentric' coincidence, since only he is in a position to view the event under a randomizing condition. The inference he alone will be able to make against this personal coincidence, in favour of a metaphysical view that will falsify it, could be called, therefore, an 'egocentric' inference.

Now, consider further that you would never have been begotten if your parents had not been begotten. And against each of them the odds once more were roughly 200,000,000 to 1. If we think of only these three required begettings, yours and your parents', we find that the probability was 1 in 200,000,000³ or 1 in 8 septillion that you would make it. And all your ancestors were similarly required for your eventual begetting. Thus each ancestral begetting was like you having won all over again at calling correctly 25 tosses of a fair coin. One wrong call in your ancestry going back to before the dinosaurs, and you never would have existed.

I say that on the basis of such utterly absurd odds against your existence on the ordinary view, you can make an egocentric statistical inference against that view and favouring the view that I have proposed. You have as your evidence that you do exist. Broadly speaking, there are only two hypotheses available to explain this evidence. There is that of a narrow prefiguration, any view like the ordinary view, in which your existence is governed by incredibly demanding conditions like those I have been discussing. And there is the hypothesis of wide or open prefiguration, in which you, through the abstract property of immediacy, are equally well present under any conditions of consciousness that the world might happen to produce. This latter would make you a sure winner, independent of sperm cell lotteries. The coincidence of your own begetting is falsified, since you *are* shown the winner, no matter what.

Here is an analogy. (The analogy, by the way, like our earlier discussion of luck, will require us to be thinking of a person according to the ordinary view. The reasoning in the analogy will be valid on the ordinary view, but

it would have to be reconstructed once we had adopted universalism.) A game is being played in an amazingly big hotel. At the start there is a drugged sleeper in each of countless rooms. The game's organizers toss a fair coin 75 times for each sleeper. Should the coin come up in a certain predesignated pattern of heads and tails, the person is awakened. Otherwise he sleeps on forever. Now the odds against calling 75 tosses of a fair coin are roughly 8 septillion to 1, the same as those against you having been produced from just your parents' and your own begettings. There are countless sleepers at the start of the game, and therefore it is enormously likely that some will awaken. But, of course, the odds against the awakening of any one sleeper considered at random are still 8 septillion to 1.

Now imagine that *you* wake up in one of these rooms and the character of the game is explained to you. You must simply count yourself unbelievably lucky. But next imagine you are told by another visitor of equal authority that the first account had been wrong. In the actual game they simply wake *everybody* up no matter what occurred in any 75 tosses of a coin. Imagine you have to choose between the two theories you have been offered.

You know, of course, that you've awakened. That's to be the evidence in the inference. There have been two hypotheses offered that might account for this evidence being what it is. One is that you have had phenomenal luck that comes to only 1 in 8 septillion. The other is that all that happened was that you won the easy game that had ensured you would be awakened. You are obliged to infer that it is 8 septillion times more likely that it was the easy game that was played, since the easy game would make it 8 septillion times more likely that you woke up.

Next let's say you are not one of the players but rather an external observer with but limited freedom. After the game has been played without your seeing it, you are allowed to choose at random the name of a player from the full list of all of them. You are then brought to only his room, where you observe through a one-way mirror that he is awake. You should then infer the enormous likelihood of the easy game having been played, since this random sample turned out as only that hypothesis would lead you to expect.

But now imagine that you know that, no matter which game is played, you are to be shown a player who awakened. If the hard game is played, you will be led to the room of one of the rare lucky winners, where you will observe him through the glass. If the easy game is played, you will be led in exactly the same style to an arbitrarily chosen player among the countless winners. When you do see the awakened player through his one-way mirror, you can conclude nothing about which game has been played. For there can be no repugnant coincidence now between a random selection and his being awake in the event of the hard game having been played. Curiously, if you overhear him voicing an inference to the conclusion of an

easy game having been played on the basis of his awakening, you can endorse his reasoning but you cannot actually share his inference. While this sample would be random relative to winning for him, it is not so for you.

Similarly, in the case of the world as it is, you observe the existence of others simply because the world has produced them. They are not for you, therefore, random samples among the potential subjects of experience. Universalism says that a random sample among these potential subjects will naturally have come into existence because all come into existence with any conscious being. The ordinary view says that only an incredibly small proportion of these potential subjects will arrive in the world. Thus a sample selected according to a criterion other than its merely existing is almost certainly not going to exist. You observe one, and only one, such random sample among potential subjects. You are tied to yourself, win or lose. You see others as winners whoever they are.

After all, back in the hotel, when you were yourself an awakened player, you were able to use yourself as a random sample, the right kind for the inference. Your observation of yourself was tied to only that one sample among potential wakers whether you awakened or not. It was as random as choosing the player in the nearest room; you *were* simply the nearest to you. Of course, you must also *win* to observe yourself. But you must be you, win or lose, and this additional requirement is what makes the sample random relative to winning. If the hard game of the ordinary view had been played, it would have been ridiculously more likely that you would have observed nothing. Only the easy game of universalism makes your self-observation understandable.

Another way of seeing the metaphysical inference is that if I am a universal subject, I *shall* be in a position to see, without prejudice, whatever organisms happen to come into existence in the begetting lottery. It will be like my being led automatically to the room of a waker in the hotel. My observation of myself as existing will not depend on any particular one of the sperm cells getting to an egg, and there will be no coincidence or luck for me in any particular organism having been produced. My observation of a winner will involve none of the unlikelihood of a randomly selected possibility among all potential organisms having been realized. For all of the specific conditions of consciousness would be accidental to the existence of the universal subject – it would exist so long as there was any experience at all. So it should be no surprise to this subject that it *does* exist.

I have mentioned parenthetically that the hotel inference would need reconstruction under universalism. Let me now describe this reconstruction very briefly. The egocentric inference disappears. If any hotel awakener would be equally I, this I could not decide between the easy and hard game on awakening in the hotel. For *I* should be awakening no matter which

game was played. Any details of organism or experiential content, since they would not be requirements for *my* awakening, would now represent no coincidence, nothing improbable for *me*. Numerically there would be but a single experience, a single awakening, despite the much greater variegation of contents on the easy game. As a practical question, however, it would be wise for me, in any of my manifestations, to behave as though the easy game had been played. For the amount of experiential content involving my being correct is likely to be increased by such a policy. If I consistently guessed that the easy game was played, I would have more experiential content of being right if I *was* right and less of being wrong if I was wrong, since, of course, I would be existing in so many more organisms when right and so many fewer when wrong. The inference *to* universalism needs no reconstruction. The rejected hypothesis in that argument is ever the unreconstructed ordinary view of a person, with its inherent egocentricity. The argument, then, remains analogous to the unreconstructed hotel inference.

Finally, a statistical argument similar to that regarding the nature of a person can be made regarding the nature of the present. If this experience of yours occurring *now* depended on objective time being at a certain point in its progress, then it would be an incredible coincidence from the perspective of this experience that it *is* now. For it would be so much more likely that this experience would now be blank, lost in the vastness of past or future, untouched by the razor-edge present. But if all experiential content simply is now by virtue of its inherent immediacy, independent of any progress in objective time, there is no such coincidence in experience being now. And so, if this, your experience, *is* now, you should infer the tremendous likelihood that this fact results from a quality in all experience rather than the fragile lucky agreement of an individual moment with a moving present time. This experience being *now*, like this experience being *mine*, is overwhelmingly likely to depend on immediacy in all experience rather than on the chance realization of incredibly narrow conditions.

IV. Changing Views

Why are we so wrong about ourselves? And why does the ordinary view hold us so strongly?

I spoke earlier about an imaginary world in which there was only one red object. I remarked on how the people there might fail to distinguish the abstract quality of red from its one observed instance. Similarly, because the abstract quality of immediacy occurs always in particular experiential contents, it is easily just identified with these.

After all, consciousness always occurs in individual organisms, at particular times. Therefore the varied content of this consciousness has in it

nothing but impressions of such individuality. The immediacy of consciousness that is the only basis for *the present* always seems centred on just one moment of time. And this same immediacy of consciousness, which is also the only basis for *the self*, always seems centred on just one organism. But these are mere appearances, illusions of perspective.

Another source of our mistake is evolution. Sexual reproduction, as opposed to reproduction by division, produces mortal organisms. In those species where such organisms survive through the possession of anything like our motivational system, they face the world with self-interest, tempered by concern for others. Their motives spring from seeing experiences as belonging exclusively to one or another individual subject. Thus future possible pains or pleasures, frustrations or successes, are seen as reasons for present action, either through self-interest or a concern for the self-interest of others. I put food in the refrigerator today for the satisfaction of myself or my family tomorrow. Creatures that regarded their futures otherwise would not be here now – unless, perhaps, a radically different basis of motivation had evolved in them. But nature seems to have taken this individualist path with humans and many other species. That this involves a metaphysical mistake, in regarding *mine* as exclusive to each organism, as bounded by that organism's complex identity conditions, detracts not at all from the survival value for such organisms of this style of seeing things. Most of their other beliefs have survival value because they tend to be true and are arrived at through discoveries of how things really are. But the view of the subject of experience as bound to an individual organism is a style of seeing things that has a survival value, for the organisms in which it is embodied, that is quite independent of whether the view happens as well to be true. Furthermore, there has been no genuine discovery on which this view could be based. Not just philosophers, but infants and many unthinking animals constantly anticipate at least their immediate futures as a series of experiences into which they, exclusively, are moving. Now, any view that owes its authority in this way to a value it possesses apart from whether it is true has yet to earn the authority of truth before a philosophical tribunal.

If we reproduced by division instead of sexually, no doubt we would have evolved with a convenient attitude towards fission. The branchings, for example, might have been regarded by the original organism simply as new people, like offspring. The original could then have prepared for the coming of these out of love. As we do not reproduce by fission, evolution has not thus readied us to think about fission. If we care about truth, we must take advantage of this lapse.

For what is exposed here, in the puzzle of fission, is a deep contradiction in the ordinary view, the one with which we *did* evolve. This view is partly right and partly wrong about the metaphysics of the subject of experience.

We see ourselves correctly as indivisible. We recognize that our existence, our presence in an experience is an all-or-nothing proposition. There can be no degrees or indeterminacy in this. For it is seen that the immediacy which marks an experience as *mine* must be *fully* in an experience if it is to be there at all. What the ordinary view does not recognize, however, is that this same immediacy must be present equally in all experience regardless of token, type or content, and is only thus truly indivisible and wholly determinate. But the ordinary view tries to make your identity depend as well on certain complex conditions of token, type, and content integration. Unfortunately, these admit of division, differences of degree, and indeterminacy. There is no way to reconcile this complexity with the simplicity of the self.

This analysis of the problem with the usual view suggests yet another argument against it, a generalized version of the earlier brain exchange thought-experiment, which might be called 'the continuum argument'. Imagine that you had been different in just the particularity of one of your atoms, or the character of one memory, or of one gene in one cell. Surely you would not have failed entirely to exist if there had been only some such slight difference in your token or type. But you would also not have been that fraction less in the world. That fractionally different body or mind would not have been a fraction less yours. You would have existed fully in all of it. The quality of being yours would have filled the experience of it.

Next imagine there having been a further slight difference, beyond that we just considered. The body or mind with the first slight difference would still have been fully yours. The reasoning, then, must apply exactly as before with this further hypothetical difference. You would have been fully present in that body and mind as well. But then we may apply this same consideration to all the slightly differing bodies and minds that might next be proposed with wider and wider differences in their tokens or types from those of your body and mind as presently constituted. And you, that subject of experience with which we started, would not have failed to be fully present in any of the continuum of all possible bodies and minds, of all conscious beings, which could finally be built from such differences. All their experience would have been fully yours. And so the experience of such *actual* beings must *be* fully yours.

Similarly, the Hindus considered how each person's individual self, his Atman, was present in such a variety of mental activities – now in feeling the sensations of the toes, now in adding two and two – and believed it was capable also of changing bodies in reincarnation. Finally, such thinking resulted in the conclusion that the 'pure experiencer' must be the same and equally present in all conscious beings, that there was but one self.

The continuum argument against the distinction between first and third

person is like a certain Newtonian argument against the distinction between terrestrial and celestial matter. Firing a cannon-ball with greater and greater velocity would increase the length of its arc over the ground till finally this would become an orbit around the earth. Thus the cannon-ball's motion would have become like that of the moon, but at no time would the changes involved have changed any of its matter from terrestrial to celestial.

Now, I am not arguing that the changes in the self are too small for us to know where to draw the line between you and the next person on the continuum, any more than the Newtonian was arguing that changes of terrestrial to celestial matter in the increasing arcs are too small for us to draw the line between the two. Neither argument, then, is a sorites argument, wherein a property gradually changed is admitted to be the property relevant to the challenged distinction. In both arguments, what is being corrected is the careless judgment that large differences in the irrelevant property can support the distinction even if small differences could not. We see the large differences are merely composed of smaller differences that do, indeed, make absolutely *no* difference towards the distinction.

The continuum argument would not work with an object such as a table – or with the body or mind itself. Recall that the earlier points about the identity of the experience and its subject in the exchange of brain quarters did not apply to the identity of the brains. Neither did the points about the novel apply to the copies. The identity of a table, body or mind is not an all-or-nothing proposition. In the later steps of changing token or type of such complex entities, we shall have to say that they are only fractionally the same as in the earlier steps or that their identity is indeterminate. In the case of a table or body or mind there is nothing like the simple abstract quality of immediacy of experience which kept *you* the same through all hypothetical differences. (Those in the world of one red object could consider red persisting across such hypothetical differences.)

I shall conclude this introduction to universalism by mentioning two important consequences that seem to follow from it.

If universalism is right, those features of our values that rest on a metaphysical distinction between persons are based on a mistake. This includes both notions of individual rights and moral responsibility. Punishment makes no retributive sense if the subject that suffers it is the same as his victim. In a corrected system of values, the ideal must be that the one subject in all his manifestations have equal concern for all sentient beings – but this would be a widened self-interested concern – altruism would be impossible. Yet this concern must include a proper appreciation of the strength of the individualist illusion in determining the happiness of this subject. He would thus still show a measure of concern for individualistic aspirations. But he would know that these were deeply mistaken.

And what does universalism tell us about the seeming finality of death?

I still believe, as I did in 1961, that if the pattern of a life is anywhere, anytime extended, that life itself is extended. But now I believe as well that there is no limit to the type of one's experience, that all experience belongs to you, to me, to the one self of experience.

Suppose you are killed on the way to visit a friend. If anywhere, any time in all reality there exists the experience of that visit, without the intervening death, you are there, making that visit. But everyone else, all conscious beings, will be there with you, as you. And you will be with them too, in everything they experience, because they *are* you.

There is no alienated self that ceases to exist when a particular organism dies. Neither is there a real progress of experiences, a moving present. And so death, when seen as an obliteration of the person and an end of his experience, is an illusion.⁵

Appendix: The Anthropic Argument

I came to the statistical argument for universalism by a curious route. In 1966, as an undergraduate, I read a piece by the philosopher F. R. Tennant about the enormous coincidence involved in the fact that the basic determinations of the physical world, determinations such as the characteristics of the fundamental forces and particles, allowed for the development of life.⁶ This coincidence is sometimes called the 'Anthropic Principle'. Here is how the physicists Carr and Rees, in their 1979 article, 'The Anthropic Principle and the Structure of the Physical World', describe this coincidence:

The possibility of life as we know it evolving in the universe depends on the values of a few basic physical constants – and is in some respects remarkably sensitive to their numerical values. Indeed, the various anthropic relations . . . in principle determine the order-of-magnitude of most of the fundamental constants of physics.⁷

In his published lecture, *Is the End in Sight for Theoretical Physics?*, Stephen Hawking describes some of these anthropic relations. He says:

For example, if the proton-neutron mass difference were not about twice the mass of the electron, one would not obtain the couple of hundred or so stable nucleides that make up the elements and are the basis of chemistry and biology. Similarly, if the gravitational mass of the proton were significantly different, one would not have had stars in which these nucleides could have been built up and, if the initial expansion of the Universe had been slightly smaller or greater, the Universe would either have collapsed before such stars could have evolved or would have expanded so rapidly that stars would never have been formed by gravitational condensation.⁸

Let's focus for a moment on just one more example of an anthropic coincidence. There are four basic forces in the universe. Two of these, the gravitational and electro-magnetic, are long-range forces, falling off in strength as the square of the distance over which they are exercised. The

remaining two, the short-range forces, are the strong and the weak nuclear force. The strong nuclear force is such that if protons come close enough together, as within the nucleus of an atom, it will hold them together. In doing so it must overcome the electro-magnetic force involved in the positive charges of the protons. Otherwise the similarly charged protons would simply repel each other as they do at greater distances. Suddenly, then, within the diameter of the nucleus of an atom, a pull overwhelms the usual repulsion of the protons. This is lucky for us, whose existence depends on the existence of atomic nuclei. But it is equally lucky for us that this pull within the nucleus is only a short-range force. It falls off in strength much more quickly than the electro-magnetic force – as the seventh power rather than as the square of the distance. If the strong nuclear force were a long-range force, it would simply pull all protons together. Without such a peculiar short-range force there would never have been positively charged nuclei with negatively charged electron shells, chemical interactions, life. By the way, this nuclear force is also obliging in one further respect. Why do the protons of the nucleus not collapse together indefinitely as the attraction increases with the shortening of the distances? The answer is that when they get too close, this same complicated force converts into a repulsion. And so the protons sit neatly in place. *Deus ex machina*.

There is an extraordinary appearance of contingency about most of these basic determinations. Consider the strength of the two long-range forces. Gravity is 10^{42} times weaker than electro-magnetic force. Think of all the different levels of strength seemingly available here. A relatively small change in the strength of either force would have made life impossible. Yet throughout the universe known to physics these determinations of the four forces and the sizes and masses of the fundamental particles remain precisely the same, fixed where life requires them. Even if there is some hidden mathematical necessity responsible for this, it must be the most amazing coincidence that this necessity agrees with the requirements of life.

The piece by F. R. Tennant that I read in 1966 argued for a religious explanation of the anthropic coincidence. A purposive Being must have shaped the basic laws. Incredible luck seemed the only alternative.

In 1968 I was writing a long paper on my ideas about death. Part of my discussion was about the size of the universe, in connection with the likelihood of the occurrence of extensions of life. Suddenly, as I wrote, it struck me that there was an explanation for the anthropic coincidence that depended on neither luck nor religion.

If there is but one physical world, and we put aside a religious explanation, it must be mere coincidence that the laws of this world are suitable for life. But imagine instead that there are also countless other physical worlds that differ widely from each other in their natural laws, their basic particles and forces. These physical worlds could be either distinct universes or

distinct cycles in the history of a single universe – one whose laws were changed every time it collapsed and exploded in a Big Bang. It could then actually be probable that at least one of these worlds would happen to fit the needs of life. But why, one might ask, should this life-producing world also be that very one that is observed? The answer seems simple; only a world that meets the requirements for the existence of observers *would* be observed. Such a way of explaining our discovery of the natural laws of our world must have enormous significance for physics, since it would mean that attempts at finding a purely mathematical explanation for the laws of our world considered in isolation must fail. The basic unifying principle of the laws that we must see would be in the requirements for the existence of us, the observers. Physicists wishing to understand reasons for the character of this world beyond the Anthropic Principle would have to explore the relationship of this world to the other worlds in the wider physical reality.

It seemed to me that this way of accounting for the Anthropic Principle could be turned into an extremely powerful statistical argument. Consider an analogy. Imagine once again that you have awakened in a strange room, from a drugged sleep. And then, as by now you might expect, it is explained to you that your waking up was the result of the way a certain game had turned out. In this story there was a fair roulette wheel with a billion numbers on it. This was spun once. If a single pre-designated number was the one that came up as a result of that spin, then you were to have been awakened. If the ball landed in any of the other of the billion numbers you would have been left forever sleeping. So here you are, awake. And all you can think is how incredibly lucky you were.

But imagine next that a second visitor denies this account. In the actual game, he explains, countless wheels were spun; and if the pre-designated number had come up on *any* of these, then you were to have been awakened. The stage is set for a statistical inference. The evidence for the inference is that you are awake. You are obliged, therefore, to infer the overwhelming likelihood of the hypothesis that had lots of wheels spinning. It seemed to me, as I wrote in 1968, that a similar inference must be made to the existence of many physical worlds.

About ten years later I learned that physicists like Hawking, Carr, and Rees were discussing such a many-worlds explanation of the Anthropic Principle (though not, as far as I have seen, in the form of a statistical argument). Then, in 1981, I discovered that both they and I had been wrong in thinking that such an explanation, as thus far described by either them or me, could work.

Back in 1974 the logician Robert Stalnaker heard me give a talk in which I presented my many-worlds argument, and he rightly resisted it. If I had won a prolonged game of Russian roulette, he pointed out, I could not

infer that lots of other games must have been played to make *my* winning more probable (though the playing of lots of games *would* have made *a* winning more probable). Not, as I say, till 1981 did I come to realize that the argument as Stalnaker heard it was incomplete. Perhaps during my earlier thinking about the Anthropic Principle I had been unconsciously depending on my view of the irrelevance of tokens to the identity of the observers of the anthropic world. Anyway, in 1981 I finally realized that unless my view about identity was joined to the hypothesis of many worlds, the latter hypothesis could not explain the anthropic coincidence. But then, as I also realized, this necessity of adopting my view about identity to explain the anthropic coincidence is itself an amazingly strong argument for accepting that view about identity along with the thesis of many worlds. Let me explain.

If we accept the many-worlds thesis along with the ordinary view of our individuation, we must see ourselves as bound up with particular tokens and thus as confined to just one of the many physical worlds – and only by luck would this be a world also capable of producing life. In fact, we would require the same luck as we did on the view that this was the only world in existence. As in Stalnaker's analogy, we would be tied to only one game of roulette.

Thus the only valid version of the explanation of the Anthropic Principle through the thesis of many worlds is a version that includes freedom from the usual insistence on the specification of a token in the identity of the subject of experience. Here, then, was my first encounter with a statistical argument for disregarding tokens. The argument I presented earlier was an extension, a generalization, of this anthropic argument. The flexibility of your token allowed you to show up wherever the other conditions for your existence had been met, including the condition that the natural laws suited the requirements for life.

To represent these points, we might add to our earlier roulette wheel analogy that there are not only countless wheels but also countless sleepers, each sleeper assigned to just one of the wheels and awakening only should the right number come up on his wheel. Under these conditions, the countless roulette wheels, though they increase enormously the likelihood of *some* awakening, in no way at all help the likelihood that *you* will awaken at *your* wheel. It would be the gambler's fallacy to think that they would be helping. But if *everyone* woke up should *any* roulette wheel come out right, in analogy to my view of the person, then it would be quite natural for you to find that you had awakened. Any success would be *yours*, and the increased number of wheels *would* help you.

We have two theories, one of which makes it incredibly unlikely that you should exist in an anthropic world, the other of which makes it natural that you do find yourself in such a world – since you will exist in any world so

long as consciousness comes about in it, and such a world must be anthropic. You must then accept the overwhelming likelihood of the theory favourable to your existence and your presence in a world that provides for it.

We can solve the puzzle of the nature of the physical laws only by adopting this theory. Thus we are obliged on statistical grounds to infer the overwhelming likelihood of the many-worlds and universalist package.

Let's briefly consider that *cyclical* many-worlds view I mentioned earlier. According to such a view, there is just one universe, but it undergoes a gravitational collapse followed by a violent expansion (a Big Bang) followed by another collapse and expansion and so on forever. In physicist John Wheeler's version of this theory, the natural laws are changed randomly in the extreme conditions at the end of each collapse. Thus each new cycle can be thought of as another physical world. And we only awaken to a cycle, a world, capable of producing life.

But this theory, of one universe in many cycles, like other many-worlds theories involving separate universes, will still not help us to explain the Anthropic Principle unless we have loosened the conditions for the identity of experience. In this cyclical theory, however, what is crucial is whether objective time differences will individuate experiences. The usual view of experience would hold that awakenings in different cycles would constitute numerically distinct experiences. Hence on the usual view it would have been far more likely, to say the least, that the identity conditions of *this* experience you are now having would have been assigned to a cycle that was not anthropic. It would, thus, have been far more likely that this particular experience had never existed. Only if this very observation, in its numerical identity, is admitted to exist whenever as a universal it arises, is its existence made more probable, and hence explained, by positing many cycles in which the universe had a chance to be anthropic and produce it. So for Wheeler's theory to work this very experience must be thought capable of existing in any anthropic success among the cycles; objective times could not individuate distinct present moments.

NOTES

- 1 Michael S. Gazzaniga and Joseph E. LeDoux, *The Integrated Mind* (New York/London: Plenum Press, 1978), pp. 3–5.
- 2 Derek Parfit, 'Personal Identity', *The Philosophical Review* 80 (January 1971), pp. 3–27.
- 3 Thomas Nagel, 'Physicalism', *The Philosophical Review* 74 (July 1965), pp. 354–5.
- 4 See Norman Geschwind, 'Anatomy and the Higher Functions of the Brain', *Boston Studies in the Philosophy of Science* 5 (Dordrecht, Holland: D. Reidel, 1966/68), esp. pp. 114–34.
- 5 I feel I must thank some of those who have shown a kind interest in this idea at various times in its long development. They have kept it going. I am grateful to Hugh Aitken, Jon Allen, Kumiko Andoh, Anita Avramides, Katherine Backhouse, Wayne Bachner, Martha Bain, Bert Barth, Julia Barwick, Alan Berger, Reverie Berlandi, Helen Betteridge, Joe Biegaj, Eleanor Bloom, Bonnie Breier, George Brennan, Gary Brodsky, Malcolm Budd, Alick Cambridge, Carol Cameron, Frank Cantor, Peter Cave, Terry Chapman, Tia Cockrell, Barry Collins, Mike Dalfen, Mark Dalrymple, Caroline Douglas-Scott, Michael Douglas-

Scott, Shona Douglas-Scott, Caroline Duthy, Penny Evans, Helen Falconer, Angela Fane, Sharon Farrelly, Robert Frankel, Maisha Frost, Tony Fruan, Rowena Gaunt, Marcus Giaquinto, Bill Gohean, Paul Goldreich, Mimi Gomis, Tish Hanifan, G. Yaël Harâr, Elsa Harford, Angela Harvey, Yvonne Harvey, Alex Hewson, Akiko Hikota, Annabelle Hoffman, Ted Honderich, Nicholas Isherwood, Hidé Ishiguro, Chieko Izumi-Holmes, Shan Jayran, Marcia Jillson, Cornel Karjohn, Hanka Kende, Helen Kim, Kyung-Sun Kim, Elayne Kirschel, Fred Kirshnit, Laurie Komonthourou, Isabel Koproski, Joel Kupperman, Dosh Kurtyanek, Martha Kurtyanek, Annie Kypriades, Stephen Lavell, Edwina Leapman, Kate Levy, Jackie Lewin, Eric Lewis, Grahame Lock, Lucy Lowles, Kenneth MacKenzie, Elizabeth Malston, Dyno Mason, Marg McArdell, Colin McGinn, Anne McMillan, Barbra Mensah, Tia Millman, Kyoko Morinaga, Catherine Morley, Ina Navazelskis, Helena Norberg-Hodge, Jane O'Grady, Alessandra Pace, Derek Parfit, Karen Percival, John Philip-Smith, Ayesha Plunkett, Nigel Quigley, Wendy Raebeck, Shahin Rafatjoo, Alison Ray, Jasia Reichardt, Tony Reichardt, Janette Ringsell, Wendy Robbins, Laney Rosenzweig, Graham Roupas, Michael Seifert, Jerome Shaffer, Amy Shapiro, Myung-Ok Shim, Susanna Sirefman, Richard Sponzo, Larry Stanton, Marek Staszkievicz, Dorothy Stein, David Stevens, Yvonne Stevens, Jerry Valberg, Elizabeth Valentine, Anne Wagstaff, Guy Warner, Nick Watley, John Watling, Gertrude Zuboff, Mark Zuboff, Nathan Zuboff, and Shoshana Zuboff.

Those who have most influenced my thinking or its presentation are G. A. Cohen, Gilbert Harman, Georg Hegel ('Sense-Certainty' in *The Phenomenology of Mind*), Thomas Nagel, Robert Stalnaker, and Peter Unger.

6 F. R. Tennant, 'Cosmic Teleology', *The Existence of God*, ed. by John Hick (New York: Macmillan, 1964), pp. 121–36.

7 B. J. Carr and M. J. Rees, 'The Anthropic Principle and the Structure of the Physical World', *Nature* 278 (April 1979), p. 612.

8 Stephen Hawking, *Is the End in Sight for Theoretical Physics?* (Cambridge: Cambridge University Press, 1980), p. 4.

Received 30 June 1989

Arnold Zuboff, Department of Philosophy, University College London, Gower Street, London WC1E 6BT, England

Methodological Individualism

Torbjörn Tännsjö
University of Stockholm

The doctrine of methodological individualism is clarified and different versions of it are distinguished. The main thesis of the article is that methodological individualism is either a false doctrine or else a doctrine compatible with functionalism, structuralism, and Marxism. Positively it is maintained that, for all we know, collective entities such as power structures may shape our beliefs and values; these beliefs and values may explain some of our actions and expectations. These actions and expectations, together with similar actions and expectations of other people, caused in a similar fashion, may in turn constitute social phenomena. This means that, for all we know, in our best explanations of some social phenomena we may well need to have recourse to collective entities.

I. Introduction

It has been argued by 'methodological individualists'¹ that some well-known and influential anthropological, sociological, and historical theories, such as functionalism, structuralism, and Marxism, can be repudiated on strictly philosophical grounds: we need not enter into difficult empirical discussions in order to see that *all* theories of these kinds are seriously flawed; the methodological 'collectivism' embraced by functionalists, structuralists, and Marxists is 'un-scientific'. Here I intend to question this argument. My main thesis is that methodological individualism is either a false doctrine, or else a doctrine compatible with functionalism, structuralism, and Marxism. This of course does not show that, in the final analysis, any of these theories is sound. It does show, however, that they should not be ruled out of court without careful examination. If they are empirically adequate, simple, and so forth, there does not seem to be any compelling *methodological* reason for not taking them seriously.

II. The Meaning of 'Methodological Individualism'

How are we to understand the doctrine of methodological individualism? Let us take as a preliminary formulation of the doctrine of methodological individualism: