

Personal Identity, Fission and Time Travel

John Wright

Published online: 10 November 2006
© Springer Science + Business Media B.V. 2006

Abstract One problem that has formed the focus of much recent discussion on personal identity is the Fission Problem. The aim of this paper is to offer a novel solution to this problem.

Keywords Personal identity · Fission · Necessary identity · Moral responsibility

The Fission Problem

The fission problem arises when a person ‘splits into two’ in such a way that –

- (a) each of the persons that is a product of the fission is a possible candidate for being the same person as the original, and
- (b) were it not for the existence of the other person that arose out of the fission, we would have little hesitation in saying that one of the products of the fission was the same person as the original.

One example of fission is given by imaginary ‘split brain’ operations. The human brain is divided into a left hemisphere and a right hemisphere. Now, suppose that each hemisphere can have its own record of memories. For example, if Smith remembers being in Los Angeles, then the memory of Smith being in Los Angeles may be stored in both his left hemisphere and his right hemisphere. Let us also suppose, as in fact seems to be case, that the left and right hemispheres can have their own ‘independent spheres of consciousness’.¹ One possible type of fission arises when we further imagine that these two severed hemispheres are placed in hitherto empty skulls of otherwise healthy human bodies. The result would be, apparently, two conscious human beings, each with their own memories of their life up until the transplanting operation.

¹For a description of some of the empirical evidence that the severing of the corpus callosum results in two independent spheres of consciousness, see R. W. Sperry, “Hemisphere deconnection and unity in conscious awareness”, *American Psychologist*, vol. 23 (pp. 723–733), especially p. 724.

The fission *problem* arises when we consider with whom, if anyone, the original person is identical after the operation. What seems undeniable, and is granted by all authors who have written about this topic, is that such an operation would give rise to *two* persons. The persons produced by the fission might have, at least for a while, identity of characteristics, but from the moment of fission they are *numerically* distinct. Now suppose A splits in to B and C. What, if any, are the identity relations between A and B and C? A first suggestion might be that B is the same person as A, *and* that C is the same person as A. This may seem initially plausible since, we are assuming, both B and C have all the same memories as A. But it is, apparently, easy to show this cannot be correct. If $A=B$ and $A=C$ it logically follows $B=C$, but we have established that $B \neq C$. Therefore, it cannot be the case that *both* $A=B$ and $A=C$. Perhaps, then A is the same person as *one of* B and C, but not the other. Let us suppose $A=B$ but $A \neq C$. But then the question arises: in virtue of what is A the same person as B, but not C? We can assume that B and C are *perfectly similar*, and that any relation B has with A, C also has with A. If so, there would be no grounds for saying that A is the same person as one of them but not the other. It seems, therefore, the only option left to us is to say that A is the same person as *neither* B nor C. But then, what happens to A in the operation? *If* A continues to exist after the operation, A must be identical to *some* person existing after the operation. The only two viable candidates are B and C. But we have seen that we cannot maintain that A is either B or C, and neither can we maintain A is both of them. The only conclusion left, then, seems to be that A ceases to exist in the operation. But this leads to some very counter-intuitive consequences! Let us suppose that, instead of B and C both surviving the operation, one of A's hemispheres – the left, say – is destroyed. A's right hemisphere survives the operation and is placed in the empty skull of another body. As a result, the person B comes in to existence but C never does. The resulting person, we may suppose, has all the memories of A. Here we would be strongly inclined to say that B is the same person as A. If this is so, then we may assert that *if* C does not survive the operation, then A will still exist after the operation; but if both B and C survive, then A will cease to exist in the operation. But would A in any way be *benefited* by C dying on the operating table? Would A be in any respect better off if C died and only B came through the operation? The answer seems to be 'No', *despite the fact that C's dying brings it about that A will survive the operation*. Therefore, if we say that if A does not survive fission, then it seems we must also say he would receive no benefit from his own continued survival. But this is a *very* counter-intuitive conclusion.

In summary, in cases of fission:

- (a) The original person A cannot be the same person as both B and C.
- (b) Neither can the original person be identical with one of B and C and not the other.
- (c) It is at least rather highly counter-intuitive to say A *ceases to exist* in the fission process, since this seems to lead us to the conclusion A receives no benefit from his own continued survival.

But, fission seems to be *possible*: split-brains may even be actual cases of fission;² yet there seems to be no way of describing cases of fission that is both coherent and plausible. This is the fission *problem*.

The aim of this paper is to offer a novel solution to the fission problem. But first we will briefly look at the existing solutions to the problem.

² See Sperry, *op cit*.

Existing Solutions to the Fission Problem

All existing solutions to the fission problem contain at least some counter-intuitive features. We will begin with what is perhaps the *least* counter intuitive of those solutions. This is the solution that has been offered by Bernard Williams.³ Williams suggests that, if a *person* is to continue to exist, it is *necessary* that their *body*, as well as their mind, continue to exist in the normal way that bodies continue to exist in the normal course of events. It is therefore a consequence of William's view that neither B nor C are the same person as A, since the transplanting of the two hemispheres into new bodies, while A's original body is left brainless, does not involve the preservation of A's body continuing to exist in a normal way. Even if B died during the operation, A would not be the same person as C since, again, A's body would not have continued to exist in the normal way. If, however, all that had happened was that *one half* of A's brain was removed from his body, while the other half remained in his body, then A would continue to exist, in his old body, after the operation. If the other hemisphere was placed in a new body, then the result would simply be a new person.

William's solution contains a feature that does not seem plausible. On this view, the continued existence of the body is *necessary* if the person is to continue to exist. This means that, for Williams, disembodied existence is not a possibility. But, *prima facie*, at least, this does not seem right. We all seem to be able to *imagine*, for example, leaving our bodies as we lie, in feeble old age, in our beds, and floating up until we reach a region in which we are surrounded by angels on clouds playing harps. This may not *actually* happen, and it may be a biological and *physical* impossibility. But on William's view, it is not even a conceptual possibility. For Williams, when (we think) we imagine ourselves without our normal material bodies, but with some 'aetherial' body, we might *think* we are imagining a conceptual possibility, but it is really no more of a conceptual possibility than that 2+2 might be 5, or that there might be a square with eight sides, or a triangle with four sides. But that this is not even a *conceptual* possibility is, *prima facie*, not very plausible.

Richard Swinburne offers a dualist solution to the fission problem.⁴ Following Descartes, Swinburne argues that the essence of mind is consciousness, and that the mind is a non-material, indivisible entity. In fission, the self goes wherever the conscious mind goes. If A splits in to B and C, only one of them will get A's mind. If it is B that gets A's mind, then it is B who, after fission, is the same person as A. We may not be able to tell whether it is B or C that gets A's mind, but for Swinburne, that is a *merely* epistemological point. What it is for A and B to be the same person is for B to get A's conscious mind, whether or not we can have any evidence that this is what has happened.

Swinburne's solution to the fission problem is simple, and, if you are drawn to dualism, attractive. But it, too, has a rather implausible consequence. Let us focus on C in the above example. If C is not the same person as A, then who is C? For Swinburne, the self – which he identifies with the conscious mind – only goes to *one* of the products of the fission – and therefore the other – in this case C – is not a person at all! C has no 'self' and is not conscious. But this leads to some very implausible consequences. The two 'entities' that are the product of fission may both be indistinguishable to an onlooker. C will have a perfectly well functioning *brain*, and C's observable behaviour may be

³ For a statement of Williams' views, see his. (1956–1957). "Personal identity and individuation." *Proceedings of the Aristotelean Society*, 57, 229–252.

⁴ One exposition of Swinburne's is given in Shoemaker, S. and Swinburne, R. G. (1984). *Personal Identity*. Oxford: Blackwell.

indistinguishable from that of B. Moreover, C's brain may store all the information that was stored in A's brain – but without consciousness – in something the same way that an unconscious computer can be said to 'store information'. C will be able to answer questions in exactly the same way that B does. Yet, on Swinburne's view, C will not be a real person at all, but a kind of automaton, devoid of consciousness, but which is indistinguishable from a real person to an outside observer. This is, I think, a possibility which it is difficult to actually accept.

Perhaps the response to the fission problem that has attracted most attention is that due to Derek Parfit.⁵ Let us return to the case of a person A about to undergo a split brain operation, in which the two hemispheres of their brain would end up in different bodies. It seemed that such a person would have nothing to gain from, for example, bribing the nurse to kill one of the new persons created on the operating table, *despite the fact that this would ensure A's own continued existence*. Parfit draws the conclusion from this that A's own continued existence is not what is ultimately in his own interests. He goes on to draw a number of conclusions in moral theory from this result. But again, the conclusion "Our own continued survival is not what is in our own ultimate interests", is to say the least, a very surprising result. Here we have a case in which what one philosopher is inclined to see as the derivation of an interesting result, another is more inclined to see as a *reductio ad absurdum* of one or more of the premises on which the argument to that conclusion is based.

Possibly the most intuitively surprising of all solutions to the fission problem is that advocated by David Lewis.⁶ What seems to be an undeniable fact is that *after* fission there are two numerically distinct persons: B and C are *numerically* distinct, although they may have identity of characteristics. If B and C are indistinguishable in all physical or material respects then we cannot, unless we accept mind–body dualism, have grounds for saying that one of B and C is identical with A while the other is not. And if we say A ceases to exist in fission we are also led to counter-intuitive results. Lewis' bold solution is to propose that prior to fission the two persons B and C were already present in the body of A. On this view, so-called cases of fission do not really involve fission, or splitting, at all – they are rather cases in which two distinct persons, temporarily occupying the same body, go their separate ways.

Lewis is not suggesting that A has a 'split-personality', or multiple personality disorder. Neither is he suggesting that there *must* already be two copies of A's memories, with one copy in the left hemisphere and another in the right, even though with human beings there is evidence that this can in fact be the case. His suggestion is that in cases of fission B and C were already present in the very same collection of molecules prior to the splitting. However, he does not hold that all bodies are occupied by two persons. It is only if, at some point in their history, that they *will* undergo a splitting, is it true that there are two persons in the one body prior to the splitting. So, presumably, every human body currently in existence contains only one person. But if there is currently in existence some human – 'Fred' – who will undergo a split brain operation in the year 2020, then the body of Fred is now occupied by two numerically distinct persons.

⁵ See Parfit's *Reasons and Persons*. (1984). Oxford: Clarendon Press. Parfit develops his views on personal identity especially on pp. 245–280.

⁶ See David Lewis "Survival and identity" in Lewis' (1983). *Philosophical Papers*, vol. 1, (pp. 55–72). Oxford: Oxford University Press. A view similar to Lewis' has recently been defended by Eugene Mills, "Dividing without reducing: Bodily fission and personal identity." in. (1993). *Mind*, vol. 102, (pp. 37–51).

Lewis' claim that one perfectly normal body can be occupied by two distinct persons is already rather counter-intuitive. But its counter-intuitiveness can be brought out even further. How many persons currently occupy my body? On Lewis' view, the correct answer is: probably only one but I do not know. If at some time in the future I am to undergo a split-brain operation, there are *now at present*, two distinct persons occupying my body. If, at some time in the future, science develops a way in which humans can divide like amoebae, and if I do divide like an amoeba, and the products of that fission in turn divide, then, on Lewis's view, my body is at present occupied by perhaps millions of distinct persons. But, whether my body is *at present* occupied by one person, or a million, depends on what scientific developments will occur in the future. All this is, I think, strongly counter-intuitive.

In summary, all existing solutions to the fission problem have at least some features that are counter-intuitive or even bizarre. The solution to be offered in this paper is no exception to this rule. But, the counter-intuitive features of the solution offered here do not constitute a reason for seeing it as any *lessgood* than the other existing solutions.

Another Possible Solution to the Fission Problem

One aspect of the fission problem on which all participants in the debate agree is that after fission takes place, B and C are *numerically distinct* persons. Although B and C may have identity of characteristics, they are now two *numerically distinct* persons. The main aim of this paper is to explore and defend the idea that B and C are *numerically identical*.

If B and C are numerically identical, then there is no obstacle to saying that $A=B$ and that $B=C$. This gives the position a pre-theoretic attractiveness. I have discovered that, when the fission problem is explained to persons who are *not* professional philosophers, such as undergraduates, the most common response is to assert, or even insist, that B is the same person as A and that C is the same person as A. It is only *after* it is pointed out that since B and C are numerically distinct, and that this leads to a contradiction, do they start to retract from this position. But the position that appears to be *pre-theoretically attractive* is that $A=B$ and that $A=C$ and that, in consequence, $B=C$.

Time Travel and Identity

But: how might it be possible for B and C to be the same person? Here we should note that there is at least one apparently conceivable means by which it is possible for there to be, what appears to an onlooker, to be two numerically distinct persons standing side by side, even though they are in fact one and the same person. That means is *time travel*. Suppose that in the year 2030 Dr Who decides to travel back in time and have a conversation with himself as he was in the year 1980. Then, to an onlooker, it will appear that there are two numerically distinct persons – one old, one young, let us say – seated in different chairs in the same room, talking to each other. From the point of the view of the onlooker, there will appear to be two persons in the room. But, in fact, the young Dr Who and the old Dr Who are numerically identical. They are the same person. Strictly speaking, when the young Dr Who and the old Dr Who are seated in the same room conversing, there is *only one person* in the room.

It is useful here to introduce the concept of a *personal world-line*. A personal world-line is the set of all points of space–time occupied by a person. We can represent the personal world line of a typical person by the line I in the following diagram (Fig. 1):

X, Y and Z are the spatial dimensions, t is time. B(P) is the persons birth at t_1 , D(P) is their death at t_2 , and the movements through the X, Y and Z dimensions represent the persons movements through space during their lifetime. The line J represents Dr who travelling back in time to have a conversation with his own earlier self. The drawing of a cube represents the room in which they have their conversation. The line segment Y represents the young Dr Who as he is seated in the room, while the line segment O represents the old Dr Who in the room. The two line segments represent distinct episodes in the life of the same person, *since they are segments in the same personal world line*. Finally, the structure K represents a person A who splits in to B and C.

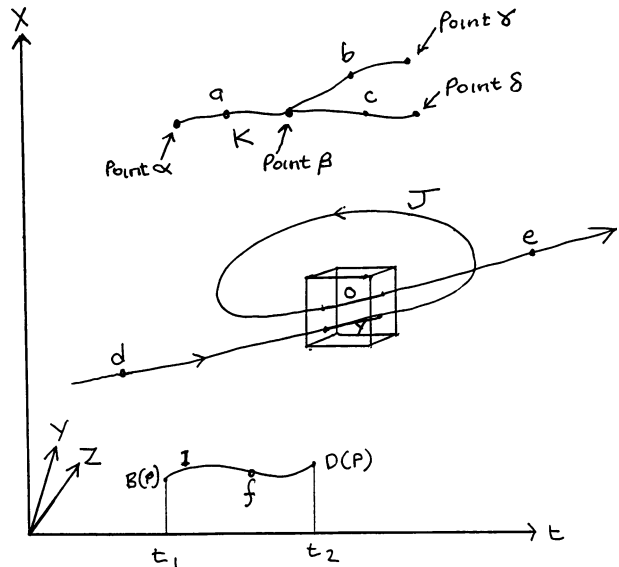
Two personal world line segments represent two episodes in the life of the same person if and only if they are different segments in the same personal-world line. This is not a philosophical thesis, it is just a consequence of our definition of a personal world-line.

It should be noted that the structure K as represented in Fig. 1 does not beg the question in favour of the thesis that $b=c$. This is because the structure K itself can be interpreted in (at least) any of the three following ways:

- (a) As a representation of one world-line; with a, b and c stages of the same person.
- (b) As a representation of three contiguous world lines; with the first extending from α to β , the second from β to γ , and the third from β to δ . On this interpretation, a, b and c are stages of three distinct persons.
- (c) As a representation of two contiguous world-lines; with one extending from α to γ , the other from β to δ . On this interpretation, a and b are stages of the same person, but b is a stage of a different person.

Clearly, to assert without argument that the first of these interpretations is correct would be to beg the question in favour of the claim that $b=c$. But, the structure K itself is

Figure 1 Representation of the personal world line of a typical person by line I.



compatible with all three of the interpretations, and so does not beg any questions in favour of the view defended here.

Referring again to Fig. 1; plainly, points d and e are on the same personal world line. Equally plainly, d and f are not. But what about a and b, and a and c? (and b and c). It is *not obvious* that b and c are on different world lines, in the way it is obvious that d and f are on different world lines. Whether b and c are counted as being on the same world line will depend on the theory of personal identity that is adopted. And there have been advanced some theories of personal identity which imply that b and a, and c and a, and therefore b and c, are on the same world line. For example, on Locke's memory criterion of personal identity, b would remember a, and so be on the same personal world-line as a, and c would remember a, and so be on the same world line as a. Therefore, Locke's criterion would lead us to say b was the same person as c, and so on the same personal world line as c. Or again, many of us would be tempted to say that a *sufficient* condition for personal identity is continuity of consciousness; that is, that if, for example, someone is continually conscious from the time at which they are lying in their hospital bed, to the time at which they pass along a dark tunnel, right through to the time at which they are surrounded by angels on clouds, then the person who was lying in the hospital bed is *the same person* as the one who saw the angels on the clouds. But it seems to be possible for fission to take place while A, B and C are all the while conscious. If so, and if continuity of consciousness is sufficient for personal identity, A would be the same person as B and the same person as C, and so B would be the same person as C. Hence, there are views of personal identity which imply that b *is*, or could be, the same person as c, and therefore on the same personal world line as c. Indeed, perhaps the reason that the fission problem has attracted so much attention is *because* so many *prima facie* plausible theories of personal identity *do* lead us to say B=C. The fission problem has attracted attention precisely because it leads us to suppose that there must be something wrong with such theories. Hence, the claim that B=C, and that b and c are therefore on the same personal world-line, is certainly *available* as a philosophical position, in that there are many theories of personal identity which imply it is true.

But of course, the claim that B=C is highly counter-intuitive and so the fission problem has been thought to discredit all such theories. The aim of this paper is to argue that it was in fact wrong for philosophers to think the fission problem discredited those theories. It will be argued that the claim that B=C is no more a conceptual impossibility than is the claim that an old Dr Who and a young Dr Who could be seated together in the same room at the same time, having a conversation. In fact, the obstacles to admitting that B=C are somewhat less than those involved in the time travel case, since there are possible generators of conceptual incoherence, present in the time travel case, which do not exist in the personal fission case. For example, if time travel is to be a conceptual possibility, it must undoubtedly be the case that it is a conceptual possibility for an effect to come before its cause. (Dr Who materialised in the room in 1980 *because* he pressed a button on his time machine in 2030.) But perhaps it is a part of our very concept of a cause that a cause cannot come after its effect. However, the requirement of backwards causation is not present in the case of personal fission.⁷

⁷ For one defence of the possibility of time-travel, see D. Lewis "The paradoxes of time travel" in, (1976) *American Philosophical Quarterly*, vol. 13, (pp. 145–52). I think it is fair to say that most subsequent discussion of the topic has tended to agree with Lewis that time travel is possible. But for a recent statement of a contrary view, see William Grey, "Troubles with time travel", (1999). *Philosophy*, vol. 74, (pp. 55–71). For a reply to Grey, see Phil Dowe, "The case for time travel", (2000), *Philosophy*, vol. 75, (pp. 441–452).

The Permissibility of Saying $B=C$

In this section various objections to the claim that $B=C$ will be considered.

Objection (1) *Saying that $B=C$ violates the principle of the indiscernability of identicals.*
According to the principle of the indiscernability of identicals:

If $X = Y$ then X and Y have all the same properties. (1)

From this principle it follows that:

If X and Y do not have exactly the same properties, then $X \neq Y$. (2)

But plainly, B and C could have different properties, and in fact are, at least after a while, extremely likely to have different properties. For example, B might grow a beard while C remains clean shaven, B might become ill, while C remains in good health, and so on. And of course, from the moment of fission, B and C will occupy different positions in space. Therefore B and C will have different properties, and so it appears that we are driven by equation (1) to say $B \neq C$.

Reply:

Equation (1) is, of course, ambiguous, and has *apparent* counter-examples. Its ambiguity lies in the word ‘have’ that appears in it – do X and Y *have* their properties at present, or do they have them merely at some time or other? A house may be painted green one year, and brown the next, and yet be the very same house. So, (1) needs to be sharpened up a little if it is to be made clear that it is not really refuted by such apparent counter examples. Perhaps the most obvious reformulation speaks not of the properties possessed by X and Y , but of the properties they possess at a time. Such a reformulation might be

If $X = Y$ then, for any time t , all of the properties that X has at t ,
 Y also has at t , and all of the properties that Y has at t , X also has at t . (3)

Alternatively, it might be claimed that it is wrong to say that the house has the property of *being brown*, or *being green*, and that the only terms which refer to permissible properties in this context expressions such as ‘being-brown-at- t ’ or ‘being-green-at- t ’.

However, reformulating (1) by referring to properties *at a time* might not be right. Consider again the case of a young Dr Who and an old Dr Who seated in the same room, conversing. They will have different properties – one may have grey hair, the other brown, one may have a beard, the other be clean shaven etc. They will have different properties at the same point in time, but they are nevertheless one and the same person. So (3) may not be quite right. A natural way of re-expressing (1) so as to overcome this possible difficulty appeals not to position in time, but to position along the world-line. One possible expression of this idea is as follows:

If $X = Y$ then the world line of X and the world line of Y are completely coincident,
and, for any point p along their world lines, at p , X and Y have all the same properties (4)

This way of reformulating (1) enables us to deal with the case of the young Dr Who and old Dr Who conversing. Although they have different properties at the same time, they do

at different points along their shared world line, and so (4) does not rule out the claim that the young Dr Who = the old Dr Who.

If time travel is a conceptual possibility, then we have a clear reason to prefer (4) to (3) as our formulation of the law of the indiscernability of identicals. But even if time travel ultimately turns out to not be a conceptual possibility, there is still a good reason to prefer (4). Equation (4) would seem to be a viable explication of the indiscernability of identicals *whether or not* time travel is a conceptual possibility. Its viability does not depend one way or the other on the possibility of time travel. On the other hand, if time travel is conceptually possible, then (3) would seem to be clearly unsatisfactory. But now, the law of the indiscernability of identicals is surely true *independently* of whether time travel is conceptually possible. Therefore, a satisfactory explication of that law should not make its truth-value dependent on the possibility or otherwise of time travel. We therefore have a clear reason to prefer formulation (4) to formulation (3).

Moreover, if we accept (4), then it is permissible to say that in cases of fission $B=C$, since although B and C may have different properties at the same time, they do so at *different* points along their world-line. The law of the indiscernability of identicals therefore does not conflict with the claim that $B=C$.

There is a related objection that it is also appropriate to discuss here. It is natural to think that if B is the same person as C, then at time t , B must be conscious of the mental states of C at that time. But if we assert the products of fission are one and the same person, we are led to deny this principle. At all points in time after fission, B will be unaware of the mental states of C. However, if we again consider the possibility of time-travel, it is clear how to reply to this objection. When the young Dr Who and the old Dr Who are seated in the same room conversing, neither is directly aware of the mental states of the other, but this does not prevent them from being one and the same person. What we should say is *not* that if $B=C$ then *at time t* they will be aware of each others mental states, but rather that they will be aware of each others mental states *at the same point on their shared world-line*. If this is accepted, then the fact that after fission B is not aware of the mental states of C is consistent with the claim that $B=C$.

Objection (2) *Saying that $B=C$ is extremely counter-intuitive.*

Perhaps the most intuitively compelling reason for denying ' $B=C$ ' is its sheer implausibility. It could be that, after their separation, B and C go on to live completely separate lives from each other, and even forget about each other's existence. For example, one of them might move to a different country. One might live in a mansion in Hollywood, the other might be a farm labourer in New South Wales. After 40 years they may have simply forgotten about each other's existence. Surely it is simply unacceptable to say two persons, living thousands of miles apart, leading completely different lives, and even ignorant of each other's existence, are in fact one and the same person.

Reply:

First we should note that something approaching this situation can exist in reality. Suppose a child is born in Peru and, while only 6 months old, is moved to Canada. The child has no memory of Peru and is never told of their early life in it. The 40 year old adult living in Canada is ignorant of the child who once lived in Peru, and the child living in Peru certainly has no knowledge of the adult who one day will be living in Canada. Yet we have no trouble in acknowledging they are the very same person.

Of course, it may be objected that in this case, the different stages of the same person are ignorant of each other at *different* times. In the fission case, we are asked to believe we are asked to believe that B could be identical to C and yet for each to be ignorant of the other

at the same time. But once again, this is no objection if we allow the conceptual possibility of time-travel. Suppose the 40 year old living in Canada steps into a time-machine and travels back in time 39 years. Then there will exist at the same time, a child in Peru and an adult in Canada, both entirely ignorant of each other's existence, but they are nevertheless the same person. The adult may stub their toe while the child will feel no pain, the child may want to be fed while at the same time the adult feels no hunger. But they are still they same person at the same time, although occupying different positions on their mutual world line.

Objection (3) *B will never become C, nor C become B.* In the Peru/Canada case, there seems to be a sense in which the child in Peru will one day 'become' the adult in Canada. But B will never become C, and neither will C ever become B.

Reply:

Strictly speaking, the child in Peru does not *become* the child in Canada, since the child in Peru is already identical with the adult in Canada. What sense there is in the idea that child becomes the adult in Canada is, I think, given by:

(DW) The adult in Canada exists *down wind on the same personal world – line* as the child in Peru.⁸

Since B is not downwind of the personal world line of C, C never 'becomes' B. Similarly B never 'becomes' C. But it does not follow that $B \neq C$, unless we add an additional claim:

(DWI) $A = B$ if and only if either A is down wind of B on a world line, or B is down wind of A.

Is DWI true? If it is true, it is not, I think, self-evident or obvious. Philosophical argument is needed to determine whether it is correct. And the main aim of this paper is to argue that there is a consistent and permissible position that entails the falsity of DWI. In the absence of any good reason for saying that DWI cannot be denied, I conclude it does not constitute an objection to the position occupied here.

Objection (4) *Saying that $B = C$ fails to do justice to the important role memory plays in personal identity.*

⁸ The notion of one part of a personal world line being downwind of another can be defined in a number of ways. Perhaps the most obvious way appeals to the idea of a 'mark'. A stream of water will carry a 'mark', such as a quantity of dye, downstream of the point at which the dye enters the water. We can use this to explicate the idea that one stage of a person is downwind of another stage. Suppose that Dr Who travels from the year 2010 to 1800. Shortly before stepping in to time machine in the year 2010, he has a large glass of whisky. When he gets out of his time machine in 1800, he is sozzled. His body has carried the mark of the whisky from 2010 to 1800. Therefore, the stage of Dr Who in 1800 is downwind of the stage of Dr Who in 2010, despite the fact that it occurs before the latter in time. Another way of explicating the idea of one stage of a person being downwind of another might appeal to the person's *subjective experience* of time: although the event of Dr Who stepping out of his time machine in 1800 occurs before the event of him stepping in to it in 2010, from the point of view of Dr Who's own subjective experience, the former event seems to occur after the latter.

Although we are assuming the adult in Canada does not in fact have any memory of being a child in Peru, it is surely *possible* for the adult in Canada to have had memories of being that child. More generally, if N is the same person as M, where N exists upwind of the same world line, it must surely be the case that, whether or not M actually remembers things that happened to N, it must surely be the case that M could have remembered things that happened to M – and that they would have, had their retention of their memories been perfect. But, no matter how perfect their memories, B will never remember the things that happened to C, and neither will C remember things that happened to B.

Reply:

This suggestion has one difficulty associated with all memory based accounts of personal identity: it is implicitly circular. What does it mean to say, eg., M *remembers* things that happened to N? Any such account must distinguish between apparent and genuine memories. If M is to be the same person as N, M must have not just apparent, but genuine memories of what happened to N. But M has a *genuine* memory of what happened to N only if M *is in fact the same person as N*. Therefore, we cannot resolve whether M has a genuine memory of things that happened to N unless we already know whether M is the same person as N. Therefore, we cannot distinguish between B's or C's genuine memories, and their apparent memories, until we already know whether or not B is the same person as C.

Objection (5) *The suggestion that B=C cannot coherently describe what happens when one of the dies and the other does not.*

Suppose that B is in an accident and dies in 2020. C, on the other hand goes on living for many years after B dies. On the account under consideration, B is the very same person as C. Therefore, since B died in 2020, and since B=C, it follows that C also died in 2020. But C was alive and well for many years after 2020. Is this not a contradiction?

Reply:

This objection is really just a special case of Objection (1). It is true that it is a consequence of the view advocated here that we must say C died in 2020 *and* that C was alive and well for many years after 2020. But this need not involve a contradiction. If time travel is a conceptual possibility then there is a way in which a person can be alive and well many years after they died: for example, if Dr Who travelled in to the distant future and then returned to his own time to die. This shows that the two claims:

- (a) X died at t_1 .
- (b) X was alive and well after t_1 .

do not necessarily involve a contradiction. There is no more of a contradiction between (a) and (b) than there is between 'Dr Who has a beard at t_1 ' and 'Dr Who does not have a beard at t_1 '. The latter does not involve a contradiction if the point on which Dr Who has a beard is different from that at which he does have a beard. Similarly, 'B died at t_1 ' and 'C was alive and well at t_1 ', together with the claim that B=C, need not entail a contradiction if the point at which B dies is a different point on their shared world line to that which C is alive and well.

Objection (6) *But isn't this view incompatible with the doctrine of the necessity of identity?*

If Cicero *is* Tully, then it is not the case that Cicero might not have been Tully. On the present account, B *is* C. If this is so, and the doctrine of the necessity of identity is accepted, then it follows that it is not the case that B might not have been C. But this is surely wrong.

The fission might never have taken place, and if it had never taken place, B would not have been C. Therefore, it is not necessarily the case that B is C. Therefore, B *isn't* C.⁹

Reply:

Let us begin by considering the Cicero/Tully case. Suppose that Cicero was given the name 'Cicero' as an infant, but did not acquire the name 'Tully' until 30 years old. But still 'Tully' refers to the same individual as 'Cicero'. If someone were to ask 'How old is Tully?' when Cicero was 31 years old, the correct answer would be 'Thirty-one years', not 'One year'. Tully was born at the very same time that Cicero was born, even though he was not at that time *called* 'Tully'. Now, consider that possible state of the world, or "way in which the world might have been", in which Cicero died when he was 25 years old. In such a world, Cicero would never have got to be called 'Tully' – no one would ever have pointed to him and said 'That is Tully'. But plainly, that world is not a world in which Cicero is not Tully, it is just a world in which he never got to be called 'Tully'. So, such a world is not a counter-example to the claim that 'Cicero is Tully' is necessarily true. Let us now apply this to the case of fission. Suppose that in the actual world Bob undergoes fission at some time *t*. After fission, one of the resulting persons is dubbed 'Fred' while the other is dubbed 'Chris'. On the view advocated here, Bob, Fred and Chris are all the same person. The names 'Bob', 'Fred' and 'Chris' all refer to the same entity. The name 'Fred', for example, does not refer just to one branch but to the whole branched structure. Fred and Chris were born at exactly the same time as was Bob; they did *not* only come in to existence when the fission took place. It is just that this person (i.e., Bob/Fred/Chris) did not get dubbed with the names 'Fred' and 'Chris' until after the fission. Now let us consider that possible state of the world in which fission never took place. This is not a state of the world in which Fred and Chris never existed. They do exist in this world: they are both Bob. But in this world the individual Bob/Fred/Chris never got to be dubbed with the names 'Fred' and 'Chris', just as Cicero/Tully never got to be called 'Tully' in that world in which he died when 25 years old. Since Bob, Fred and Chris are all the same person in this possible world, it does not constitute a counter-example to the claim that 'Fred is Chris' is necessarily true. Hence, the account offered here *is* compatible with the doctrine of the necessity of identity.

Objection (7) *But doesn't this view create problems for our notions of moral responsibility, and culpability?*

It would be absurd to *blame* B for something done by C, or to blame C for something done by B. Yet on this view they are the very same person and so *ought* to be held responsible.

Reply:

It does not follow at all that this account leads to counter-intuitive claims concerning moral responsibility. Suppose a child alive in 1940 and an adult alive in 1980 are the same person. It would, of course, be absurd to hold the child responsible for some crime committed by the adult in 1980. So, it might be suggested that a person can only be held responsible at a time *t* for some crime if time *t* is *after* the time of committing that crime. But if time-travel is a conceptual possibility, that won't quite do. Suppose a scientist with a time machine travels back in time and commits a crime. At the same time that the scientist commits the crime he is also a 10 year old boy. It would plainly be absurd, one year later, to hold the 11 year old boy that is the same person as the scientist responsible for the crime.

⁹ This type of difficulty was raised by Robert C. Coburn in his article "Personal identity revisited" in, (1985), *Canadian Journal of Philosophy*, vol. 15 (pp. 379–403), especially pp. 386–387.

This suggests we need an alternative account of the conditions under which a person can be held morally responsible for an action:

A person at point p_1 on their personal world line can only be held responsible for an action committed by them at point p_2 on their personal world line if p_1 is downwind of p_2 (5)

Note that this is only a necessary condition for responsibility, not a sufficient condition. Let us now apply this principle to the case of fission and B and C. As we have already noted, B is not downwind of C, and neither is C downwind of B. You must travel up wind along the personal world line, for at least some distance, to reach B from C and C from B. Therefore, on the equation (5), it is entirely appropriate for B to not be held responsible for the actions of C, and for C to not be held responsible for the actions of B.

There is one respect in which the account offered here can deal very simply and naturally with what we intuitively feel to be instances of culpability in cases of fission. Suppose that before fission, A commits some crime. A then goes into the laboratory of a scientist with a ‘personal fission device’. A splits into B and C, where B and C are both perfect copies of A, complete with memories and personality characteristics. I think we intuitively feel that both B and C ought to be held responsible for, and be punished for, the crime committed by A. And on the view advocated here, it is very natural why this should be so. On this view, B and C are both the very same person as A, and are both down wind of A, and so can be praised and blamed for things done by A.

Objection (8) *But the account offered here cannot deal with cases of fusion in a way that is remotely plausible.*

Suppose D and E fuse in to person F. If cases of fusion are to be dealt with in the same way as cases of fission, then we must say that D and E are the very same person. But this is surely an unacceptable consequence. D and E might be born in different locations, and, up until the point of fusion, be completely ignorant of each other’s existence. It is surely absurd to say they are one and the same person.

Reply:

The arguments that were used to defend the claim that B and C are the same person after fission can also be used to defend the claim that D and E are the same person before fusion. D and E have different properties at the same time, but this is compatible with D being identical with E if we accept (4) as our formulation of the principle of the indiscernability of identicals. D and E can be ignorant of each other’s existence, and unable to recognise each other. But so can B and C, and so can the time travelling Canadian adult and the Peruvian child. Just as B can be alive after C’s death even though $B=C$, so can D be alive prior to E’s birth even though $D=E$. Finally, E cannot be held responsible for things done by D, since E is not downwind, on the personal world line of D.

It may be doubted whether the account offered here squares with our intuitions about responsibility. Suppose D commits a crime. F is downwind of D, and so it might be thought that on the account offered here F will be able to be held responsible for the crime committed by D. But this does not seem quite right, because E is *also* the same person as F. If F is punished, so is E. But this does not seem right. Intuitively, we feel that in some sense it is only ‘that part’ of F that has its origins in the D ‘root’ that should be punished, and that the part of F that had its origins in the E ‘root’ should go unpunished. In fact, in merging with E, the criminal D is engaging in some thing like an extreme form of hostage-taking. It is as if D could say: “You cannot punish me now without also punishing the innocent ‘E’

part of me.” And our intuitions surely are that it would be wrong to punish the innocent ‘E parts’ of the person.

However, it is not really a consequence of the view offered here that it is appropriate to punish F for a crime committed by D. Although being downwind of D on the world-line is a necessary condition for culpability, it is surely at least very far from clear that it is a sufficient condition. Suppose a person commits some crime in their youth. By the time they have reached old age, not only have they forgotten about this crime, they have totally rejected the values that led them to commit it. Although the old person is down-wind of the crime on the world-line, it is very far from clear that it is appropriate to punish the old person for the crime. So, the claim that being down-wind on the world-line is a sufficient condition for culpability does not seem to be correct. We are, therefore, able to maintain that in the case involving D, E and F, a sufficient condition for F’s culpability does not obtain.

Parfit’s Physics Exam

The view given here can give a natural account of a perplexing case imagined by Derek Parfit in his ‘Reasons and Persons’. Parfit imagines a situation in which he only has a short period of time in which to complete a question in a physics examination.¹⁰ There seem to be two possible ways of solving the problem, but he does not have time to explore both. So he divides his brain into two independent spheres of consciousness, and gets his left hemisphere (and right hand) to work on one approach while using his right hemisphere (and left hand) to work on the other. He then re-unites his two hemispheres into a single domain of consciousness, and sees which of the two approaches has worked the better.

This is, of course, a variant on the fission problem; but it has a number of differences. First, it seems very hard to deny that the (united) Parfit before fission is the same Parfit as the (re-united) Parfit after the subsequent fusion. But what are we to say of the two sub-Parfit’s that temporarily exist while his consciousness is divided? Parfit argues that it is not plausible to claim that there are two new persons, who both have a temporary existence, and who are replaced by the ‘re-born’ Parfit when he re-unites his consciousness. But on the view advocated here, we do not need to say this. We can say that the person tackling the physics problem by one approach, and the person tackling it by the other, are one and the same person, namely: Parfit.

Conclusions

In this paper it has been suggested that the fission problem can be solved if it is asserted that the (apparently) distinct persons that are created in fission are in fact one and the same person. The suggestion has been defended against objections. Of course, the proposed solution has its counter-intuitive features, but in this respect it is no different from other solutions that have been offered. The proposed solution deserves to take its place beside other extant solutions to the problem.

¹⁰ See Parfit’s *Reasons and Persons*, pp. 246–248.