TRENTON MERRICKS

FISSION AND PERSONAL IDENTITY OVER TIME[*]

Philosophical Studies 88 (1997): 163-186.

It seems possible for persons to undergo "fission". Perhaps half of one's brain can be kept alive and placed in one body, and the other half of one's brain kept alive and placed in another body. Derek Parfit imagines this occurring to him, and wonders:

> What happens to me? There seem only three possibilities: ...I do not survive; ...I survive as one of the two people; ...I survive as both [in that I have two bodies and a divided mind]. (Parfit, 1971, 5).

Parfit thinks that we should reject each of these three possibilities. Furthermore, he wants to deny

> ...that to any question about personal identity, in any describable case, there must be a true answer. For those with this belief, [fission] is doubly perplexing. If all the possible answers are implausible, it is hard to decide which of them is true, and hard even to keep the belief that one of them must be true. If we give up this belief, as I think we should, these problems disappear. (Parfit, 1971, 8).

Parfit's primary use of the fission argument is to motivate the claim that identity does not matter in survival. But I think he also holds that in a case of fission we have reason to believe that there is no true answer to the question of personal identity, no determinate fact of the matter at all. What else could he mean when he suggests that we give up the belief that "to any question about personal identity, in any describable case, there must be a true answer"?[1] Moreover, even if Parfit does not intend his discussion of fission to constitute an attack on determinate identity, such an attack is surely suggested by claims Parfit has made. And these claims have been developed into full-fledged

arguments against determinate identity; that is, against the thesis that, necessarily, there is a determinate answer to every question of personal identity (cf. Johnston, 1989).[2]

Anyone who has read Lewis (1976) should know that, given four-dimensionalism, fission poses no threat to determinate identity.[3] However, there is no similar defense of determinate personal identity in the face of fission given explicitly endurantist assumptions.[4] The endurantist rejects four-dimensionalism with respect to putatively dividing persons; she thinks that three-dimensional persons can persist from one time to another and that identity over time is simply numerical identity.[5] Since the endurantist thinks that identity over time just is identity, if she is committed to determinate identity, then she is committed to determinate identity over time. That is, she is committed to the thesis that, necessarily, all claims to the effect that P at t is identical with P* at t* are determinately true or determinately false.[6]

Throughout this paper, I will assume that persons endure.[7] While my primary aim is to defend the claim that fission does not threaten determinate identity, a secondary aim is to defend endurance. For if the possibility of persons undergoing fission were to show enduring persons inconsistent with determinate identity, this could easily lead one to reject endurance, rather than determinate identity. I will argue, in fact, that not only is the possibility of personal fission consistent with determinate identity, but that (given endurance) fission does not even render implausible the claim that, necessarily, there are determinate facts of the matter about personal identity over time.

How might one move from a case of fission to the rejection of determinate personal identity?[8] In other words, how is the "fission argument" supposed to go? One could follow Parfit's lead and argue that fission leaves only implausible options for the friend of determinate identity, and that therefore determinate identity is implausible and ought to be rejected. One might even think that Parfit's argument not only suggests a reason for finding determinate identity implausible, but that it contains the seeds of an outright refutation of the view.[9] Let us consider first whether there is a Parfit-style

argument for the claim that fission and determinate identity lead, not to mere implausibility, but to impossibility. In the concluding portion of the paper I will ask whether fission, if not proving determinate identity false, shows determinate identity to be implausible.

FISSION AND THE POSSIBILITY OF DETERMINATE PERSONAL IDENTITY OVER TIME

So assume for reductio that there must be a determinate fact about a person's identity over time. Add as a premise that a case of personal fission resulting from placing each hemisphere of a person's brain into a different body is possible.[10] Let's name the pre-fission person 'P'. The putative person whose brain was the left hemisphere of P's brain we shall name 'L', and we shall call his right-hemisphered (more creative?) counterpart 'R'. From this we get the supposedly unacceptable consequence that possibly:

(1) No one existing after fission is identical with the pre-fission person P

or

(2) Either post-fission L, or the post-fission R (but, of course, not both), is identical with the pre-fission person P

or

(3) L and R jointly compose a scattered person who is identical with the pre-fission person P

There is, of course, logical space for options other than (1), (2) or (3) (even given a case of fission, the denial of no two of these implies the third); however, these are the three most plausible options. So let's grant, for the sake of argument, that determinate personal identity combined with fission entails the disjunction of (1) through (3).

If the disjunctive ((1) or (2) or (3)) is not possibly (determinately) true, and if fission is possible, then the reductio is successful and we have a proof that determinate identity must be rejected. The only strategy I know of for discrediting ((1) or (2) or (3))

is to attempt to discredit each of the individual disjuncts. I will focus on attacks on (1) and (2), but not (3). Parfit would reject (3) because it entails some real oddities.[11] But the occurrence of fission itself is a real oddity, so I'm not certain that we should take the fact that one oddity entails another to show (3) to be impossible. And perhaps the friend of determinate personal identity should endorse the possibility of (3) if the arguments against (1) and (2) are successful. I hope to show, however, that the arguments against (1) and (2) fail.

To repeat, in this section of the paper I will attempt to block arguments for the conclusion that ((1) or (2) or (3)) is impossible. The plausibility of ((1) or (2) or (3)) is the subject of the subsequent section.

Option (2) Considered

Although Parfit is interested in implausibility, we can, I think, find in Parfit an intuitive reason that one might judge (2) impossible:

> The trouble here is that...each half of my brain is exactly similar, and so, to start with, is each resulting person. So how can I survive as only one of the two people? What can make me one of them rather than the other? (Parfit, 1971, 5).[12]

I think that the question—what can make P identical with (say) L rather than R?—would be an odd one if 'L' and 'R' were rigid designators; it would be akin to asking what "makes" P identical with P, rather than some object other than P. But, if 'L' and 'R' are definite descriptions, this question doesn't seem quite so odd (compare: What makes Bach (identical with) your favorite composer?). This is the first reason—others will be addressed below—that the fission argument is most charitably interpreted by taking 'L' and 'R' as non-rigid definite descriptions.[13]

The answer to the rhetorical "what can make me one of them rather than the other" and "what can make P identical with L rather than R?" is presumably supposed to be "nothing". The defender of the fission argument must claim—in rejecting (2)—that

neither L nor R possesses any property the having of which is sufficient for (or "makes") its identity over time with P, the pre-fission person. And, of course, that L and R each has no property which "makes" identity with P entails that (2) is not true only if having some such property or other is necessary for the identity of a post-fission person with P.[14]

So the argument against (2) rests on three assumptions. The first is, obviously, that there are some properties or features such that, necessarily, if one of them were possessed by one of the post-fission persons, then that person would be the pre-fission person. The second is that having at least one or another of these properties or features is necessary for identity with the pre-fission person. The third is that the post-fission persons are alike in lacking all the relevant properties—including, crucially, whatever properties would satisfy the demands of the first two assumptions.

How plausible is the first assumption, the assumption that there is some property, the having of which would entail the identity of, say, L with the pre-fission person P? If the list of candidate properties is unrestricted, then the assumption is unquestionably true. A few examples of such properties: L's being such that he is identical with P; L's exemplifying an individual essence of P (e.g., the property of being P); L's being such that immediately before fission he was (the only object) located in the exact place occupied by P immediately before fission; and even some weaker claims that might not entail the identity would probably be enough to guarantee, in some sufficiently strong sense, the identity—for instance, L's being such that he existed before fission.

But we cannot build into a case of fission the claim that L and R are exactly similar with respect to lacking all of the above sorts of properties. Suppose we said that both L and R were alike in failing to exemplify an individual essence of P. This is tantamount to building into a case of fission the claim that the pre-fission person, P, is, after fission, identical with neither L nor R. And this, of course, begs the question if one is trying to argue from a case of fission to the conclusion that (2) is not possibly true. So just as one cannot stipulate of a case of fission that (2) is false, so one cannot stipulate of

a case of fission that the resultant persons are exactly similar with respect to lacking all of the features that would obviously satisfy the first assumption—features which clearly and uncontroversially would "make" one of the resultant persons identical with the pre-fission person.

No foe of determinate identity provides a collection of features such that: A) necessarily, a resultant person is identical with the pre-fission person if and only if she or he exemplifies at least one or another of the features in this collection, and B) one may claim—without assuming that (2) is not true—that both resultant persons lack every feature in this collection.[15] Until someone fills in these essential details surrounding exact similarity, the fission argument is incomplete and any objection to (2) by way of exact similarity of resultants is a mere sketch of how an objection might go.

No defender of the fission argument has shown that (2) cannot possibly describe the outcome of a case of fission. No defender of the fission argument, therefore, has shown that ((1) or (2) or (3)) is not possibly determinately true. It might seem, then, that there is no reason to examine attacks on the possibility of (1). However, the argument against (1), if successful, would prove not only that (1) is not possible, but that (regardless of the possibility of (2)) determinate identity results in contradiction.

Option (1) Considered

One of the possibilities countenanced by the friend of determinate identity is that when a person undergoes fission:

(1) No one existing after fission is identical with the pre-fission person P.

Again, in Parfit we find the beginnings of an objection to the possibility of (1). He says:

> We agreed that I could survive if my brain were successfully transplanted. And people have in fact survived with half their brains destroyed. It seems to follow that I could survive if half of my brain were successfully transplanted and the other half were destroyed. But if this is so, how could

> I not survive if the other half were also successfully transplanted? How could a double success be a failure? (Parfit, 1971, 5).

Imagine that the left half of P's brain is destroyed, and the right half is successfully transplanted into a new body; call the person—whomever he may be—who results from the transplant 'R' (so note again that 'R', like 'L' below, is a definite description; more on this to follow). Let us agree, at least for the sake of argument, that after the transplant, R would be P. Let's call the situation in which this occurs 'C'. So:

(4) If R were in situation C, R would be P.[16]

An analogous story could be told in which the right half of P's brain is destroyed and the left transplanted, and P again survives. Let's use 'L' to refer to the person who results from this transplant, and assume that in such a case L would be P. So with L and some situation C*, it seems that:

(5) If L were in situation C*, L would be P.

How does fission enter the argument? Fission is puzzling precisely because it involves two processes, either of which alone would provide for the continued existence of P, occurring jointly. So the possibility of fission shows that R's being in situation C and L's being in situation C* could possibly co-occur.[17] So we can add:

(6) It is possible that both R be in situation C and L be in situation C*.

Note that if (6) is true, we learn something surprising about situations C and C*. Although C is originally described as a situation that occurs when the left half of P's brain is destroyed, this destruction cannot be essential to C's occurring, since this would rule out (6). Likewise with C* and the destruction of the right half. (6) is of course integral to any argument from "double success". I think it is best, therefore, to understand C as something like a diachronic arrangement of the atoms (or cells) on the right side of P's brain that sustains a mental life (thus making no essential reference at all to what occurs on the left side). Which diachronic arrangement of atoms is included C? The one that would occur if P underwent right half brain transplant with left half

destruction. This is of course consistent with asserting C could occur absent left half destruction. Similar remarks apply to C*. As we shall see below (when discussing "The Process Principle"), this understanding of C and C* is amenable to the strongest case that can be made against determinate identity by way of fission.[18]

What is supposed to follow from (4) through (6)? Perhaps we are to conclude:

(7) If R were in situation C and L were in situation C*, then R would be P and L would be P.

Note that if we reject (7), it is hard to see why the co-occurrence of C and C* should be called a "double success". I think, therefore, that we should understand the proponent of the fission argument as thinking that something like (7) follows from a case of fission. From (6) and (7) it follows that

(8) It is possible that both L is P and R is P.

But L and R are distinct and (8) is necessarily false.[19] If successful, this argument would show that our defense of (2) does nothing to help the friend of determinate identity, since we can deduce (8) from a case of fission.[20] If the possibility of personal fission entails (8), then we must reject the claim that possibly dividing persons enjoy determinate identity.[21] But the argument that (8) follows from a case of fission is invalid. Consider the crucial moves of this argument:

(4) If R were in situation C, R would be P.

(5) If L were in situation C*, L would be P.

(6) It is possible that both R be in situation C and L be in situation C*.

Therefore,

(7) If R were in situation C and L were in situation C*, then R would be P and L would be P.

The argument is of the following form:

p subjunctively implies q

r subjunctively implies s

It is possible that both p and r

Therefore,

(p and r) subjunctively implies (q and s)

Suppose: p=Murphy is taking a shower, q=Murphy is singing, r=Murphy is on television, and s=Murphy is calmly giving the weather report. Let us suppose that Murphy is a musical and modest meteorologist, and that were he in the shower, he would be singing, and were he on television, he would be calmly giving the weather report. It doesn't follow from these facts and the possibility of Murphy's showering being televised, however, that were he taking a shower on television, he would be singing and serenely reporting on the weather.

How should the defender of the fission argument respond to this charge of invalidity? The above argument schema could be made valid by adding that p and r not only subjunctively imply, respectively, q and s, but that p and r entail q and s. So one way to make the argument from fission valid is to assert:

(4*) R's being in C entails that R is identical with P

and

(5*) L's being in C* entails that L is identical with P

There are other ways, of course, that one could make this argument valid. For instance, one could add as a premise that (4) entails (7). But apart from affirming (4*) and (5*), there is, I think, no plausible and non-question begging way to render valid the argument from fission and determinate identity to (8). Moreover, as we shall below, the claim that (4) and (5) entail (4*) and (5*) underlies the strongest version of the fission argument actually defended.

If one rejects (4*) and (5*), yet endorses (4) and (5), then one, presumably, accepts that whether R would be P could depend on things that happen not only to R, but also to those parts on the left-hand side of P's brain. So it may be that R would be P if the left half of P's brain were destroyed, but, if not, then R would not be P. But

endorsing such a possibility, one might object, is unacceptable because it involves rejecting what David Wiggins calls "The Only a and b Principle". This principle states that facts about objects distinct from a and b are irrelevant to whether a is identical with b (Wiggins, 1980, 96). So, one might conclude, (4) and (5) combined with this principle seem to imply (4*) and (5*).

I think that the Only a and b Principle is true (perhaps trivially so) if we restrict 'a' and 'b' to rigid designators. But we should not take this principle to be true if we allow 'a' or 'b' to include (non-rigid) definite descriptions, because then it would clearly be false. Imagine that 'b' is the definite description 'the tallest woman in town'. Whether a is identical with b, that is, whether a is the tallest woman in town, obviously depends on facts about objects other than a and b; it depends on the fact that c, who is taller than a, is out of town. Likewise, if 'R' and 'L' are definite descriptions, whether P is identical with L might depend on the fact that R is not in C.[22]

And 'R' and 'L' must be definite descriptions if both (4) and (5) are to be possibly true. If 'R' and 'L' were rigid designators, it would be absurd to make claims like: were such and such the case, R (and not L) would be identical with P (i.e., (4)), but if instead thus and such occurred, then L (and not R) would be identical with P (i.e., (5)). This would violate the necessity of identity. (Contrast this with similar claims made using non-rigid definite descriptions: e.g., Plato was not identical with the most famous Greek to drink hemlock, but he would have been had he, but not Socrates, drunk hemlock.) Furthermore, if 'R' were a rigid designator that referred to P, we could not legitimately assume that R (i.e., P) exists after fission, because whether it is determinately true that P exists after fission is one of the points of disagreement between friends and foes of determinate identity.

But a number of philosophers have endorsed a principle—some even arguing that it is the correct interpretation of Wiggins's the Only a and b Principle[23]—that would result in (4) and (5) implying (4*) and (5*). This principle, which I will call 'The Process

Principle', involves the claim that if some process (such as a particular diachronic arrangement of the atoms from the left hemisphere of P's brain) secures a person's continued existence, then that process, in virtue of its intrinsic features, necessarily secures that person's continued existence. Thus Mark Johnston, in arguing that fission "is a case in which there are no determinate facts of identity" (1989, 371), claims:

> If in one possible world w a process p secures the survival of a person x then in any world w' in which p occurs and is intrinsically exactly as it is in w, in that world w' p secures the survival of x. (1989, 381)

Johnston gives us some insight into what is intrinsic to such a process with:

> Roughly, intuitive conceptions of the nature of the total process which secures a given person's survival will locate that process wholly within the space-time envelope or four dimensional total position swept out by that person's body over his lifetime. (1989, 381)

What is intrinsic to such a process goes hand-in-hand with the space and time occupied by a person's body; so any such process that guarantees a person's continued existence will also guarantee that that person is a certain size. The relevance of this observation will be apparent below.

The defender of the argument from "double success" will presumably claim that it is in virtue of the intuitively intrinsic processes found in C and C*—processes involving, e.g., how the proper parts from of one side of the brain are interrelated and functioning together to sustain a mental life—that each would secure the survival of P when occurring in isolation. So I think that if the Process Principle were true, one could insist that (4) and (5) do entail (4*) and (5*), and thus that the fission argument can be rendered valid.

But I think the Process Principle should be rejected because it has unacceptable results, and not just in cases of fission. Consider a case, not of fission, but of transplant of the left half of P's brain and the destruction of the right half. Of course, the defender

of the fission argument must assume (5) is true and that, in such a case, P would survive as the recipient of the left half of the pre-fission brain. Now consider the process (call it 'C**') that occurs during the first segment of this operation. C** commences when the left half of P's brain is removed from his body, and ends some time before the left half has been successfully transplanted into the new body. At the conclusion of C**—after removal, before transplant—what parts does P have?[24] It seems that P is composed of only those parts which were on the left side of P's brain before fission. P's predicament is better described, so it would seem, as body and right half brain amputation followed by body transplant.

Now, if the Process Principle is true and C** counts as a "survival securing process", C**'s occurring results, necessarily, in P's being hemisphere-sized, having just those parts that were originally on the left side of his brain. But C**, even though initially described as a case in which the left half of P's brain is removed from the skull, could possibly occur while the left half of P's brain remains encased in his skull along with the right half; this is possible since, for example, not being in a skull is not an intrinsic feature of C**. All that is essential to C**'s occurring supervenes on certain relations holding between the atoms (or other proper parts) that formerly composed the left half of P's brain, and those atoms exemplifying certain intrinsic properties (for that is all that is contained within the space and time swept out by P's diminutive "body" in C**). What we have here is strictly analogous to initially describing C as a process that involves the transplant of the right hemisphere and destruction of the left, but then going on to insist that C could occur, in a case of fission, absent any destruction of the left hemisphere.

Suppose the Process Principle is true and that C** occurs while the left side of P's brain remains encased in P's original skull alongside its right counterpart. From this it follows, necessarily, that P shrinks to hemisphere size! Again, if the Process Principle is true, that the atoms (and so on) on the left side of P's brain are arranged in the ways they

would be were the left hemisphere removed entails, not merely that P would have a divided consciousness or a blackout or a loss of motor control, but that P would shrink to the size of a brain hemisphere—even if both the right and left halves of P's brain remain in P's original skull.

But I think this claim of entailment is mistaken; it seems possible that the atoms on the left-hand side of P's brain act in funny ways but P not shrink, especially if the rest of P's brain and body function more or less normally. So I think we should reject the Process Principle.[25] Nothing, then, stands in the way of rejecting (4*) and (5*), even if we accept (4) and (5).[26] Thus we have no reason to think that the argument from double success is valid.

I'd like to address, briefly, a familiar issue raised by the possibility of P having only those parts which at one time compose a hemisphere of his brain. Suppose that among P's parts is a bona fide object that is a right hemisphere. Now suppose that, as it would seem, that object is removed. The hemisphere-shaped object sitting outside the brain, awaiting transplant, is presumably identical with the right hemisphere formerly in P's brain. But if that is so, it cannot be P, because it was a proper part of P (a proper part of P is distinct from P, therefore it cannot be (now or ever) identical with P).

If one wants to endorse (4) (or (5)), then, given endurance, I think one must deny that there are such objects as hemispheres of the brain. Of course, one who makes this denial has the not-too-difficult task of describing a case of fission in a way that does not involve objects like brain hemispheres or halves; this could be done by discussing, for instance, the atoms that lie to one side or the other of the center of the brain. Alternatively, one might maintain that there are brain hemispheres, but claim that upon hemisphere removal, P becomes wholly co-located (but not identical with) with a brain half that was once a proper part of it. That, too, would be surprising.

Insofar as one is troubled by non-existent hemispheres, persons who shrink to half the size of a brain, and wholly co-located distinct entities, one has reason, not to deny

determinate identity, but to deny (4), (5), (4*), and (5*).[27] One has reason, that is, to reject the premises crucial to the argument from "double success".

We have no compelling reason to think that fission combined with determinate identity results in contradiction or necessary falsehood of any sort. Option (1) has not been shown to be necessarily false, nor have we found an argument against the possibility that an object survive a case of fission as one of the two fission products. The fission argument has failed to prove that determinate identity, when faced with fission, entails the impossible.

FISSION AND THE PLAUSIBILITY OF DETERMINATE PERSONAL IDENTITY OVER TIME

Suppose the defender of the fission argument claimed that although it is possible that there be a determinate fact of the matter as to the identity over time of the dividing person, this is implausible. She might grant that the friend of determinate identity can provide logically consistent interpretations of cases of fission, while nevertheless insisting that such interpretations are implausible, perhaps deeply so. Therefore, she might conclude, determinate identity is itself implausible. This challenge, we should recall, is closer to the one Parfit actually raises.

Before directly addressing this challenge, two points need to be made. First of all, I will presuppose, as does the challenge itself, only an intuitive understanding of implausibility. Nothing I say below turns on any particular analysis of that concept.

Secondly, even if we grant that fission shows that determinate identity entails the implausible, this gives us good reason to abandon determinate identity only if a more plausible option is open to us. But it is not obvious that the rivals of determinate identity are themselves plausible. Some even think that the denial of determinate identity leads to absurdity (see Evans, 1978). So even if fission shows that the friend of determinate identity is committed to implausibility, it is far from obvious that we can free ourselves from all implausibility simply by abandoning determinate identity.[28]

But is determinate identity committed to implausibility? As we saw above, determinate identity, combined with a case of personal fission, leads to one of three possible results:

(1) No one existing after fission is identical with the pre-fission person P

or

(2) Either post-fission L, or the post-fission R (but, of course, not both), is identical with the pre-fission person P

or

(3) L and R jointly compose a scattered person who is identical with the pre-fission person P

If Parfit (1971, 8) is right, then each of these possible outcomes of fission is implausible. So, one might conclude, determinate identity entails the implausible and is, therefore, itself implausible.

There are two extremely serious problems with this argument, either of which alone would be enough to discredit it thoroughly.

First of all, suppose we grant that each of the three possible outcomes of a case of fission, given determinate identity, is implausible (this is, of course, to grant a great deal). It does not follow from this, however, that determinate identity entails the implausible in the face of fission. For determinate identity combined with P's fission does not entail (1); it does not entail (2); and it does not entail (3). Rather, determinate identity combined with P's fission entails the following three part disjunction: ((1) or (2) or (3)).

One might argue that if each of (1), (2) and (3) is implausible, then so is the disjunctive ((1) or (2) or (3)). But this inference is clearly fallacious. Consider an adaptation of a well-worn example: Persons P_1 through P_{1000} buy tickets in a thousand ticket lottery. It is implausible that P_1 will win, implausible that P_2 will win, and so on. Yet, the following disjunction is not implausible: P_1 will win or P_2 will win or...P_{1000} will win.

Or take a different sort of example. Suppose we have an argument with premises, the denial of each of which is implausible, but a conclusion we know to be false. (This is one way to characterize a paradox.) The denial of each premise is implausible, but because we know that the conclusion is false, the disjunction of the denials of each of the premises is not implausible. In such a case, this disjunction is just as plausible as the denial of the conclusion. As an illustration, consider the following Sorites paradox:

(8) Any municipality with over seven million citizens is a very large city.

(9) If a municipality is a very large city, then it remains a very large city if it loses exactly one citizen.

Therefore

(10) A municipality of one person is a very large city.

This argument is valid. The conclusion is false. Many will find the denial of (8) implausible and they will also find the denial of (9) implausible. Note that if the denial of (9) is true, then there is a particular number of persons that is the least number of persons that can be had by a very large city, even though, presumably, none of us knows what that number is. So, many may find the denial of (8) and the denial of (9) both implausible. But many of those same persons, in recognizing the falsity of (10), will not find the denial of (8) disjoined with the denial of (9) implausible; they will be sure this disjunction is true. And there is nothing irrational in this position.

Consider a third sort of example of implausible disjuncts but not implausible disjunction. One might find the claim that individual thoughts are identical with particular brain states implausible. And one might find the claim that a thought is identical with a "non-physical" state implausible. And one might find implausible the claim that there are no thoughts. In fact, one could quite rationally find each particular theory along these lines implausible, yet not find implausible the disjunction of all the theories.

I think that the fact that a disjunction of implausibilities need not itself be implausible follows, more or less, from a close cousin of the "paradox" of the preface—this "paradox" is that one may rationally believe a conjunction to be false while believing all of the conjuncts.[29] The "close cousin" of the paradox I am interested in has to do with plausibility. So suppose I think each of not-a, not-b, and not-c is plausible, but do not think that their conjunction, not-a & not-b & not-c, is plausible. Add that, because of this, I find each of a, b, and c implausible, but do not find implausible the denial of not-a & not-b & not-c, which denial is, of course, a or b or c.[30]

So I contend that, at the very worst, determinate identity combined with fission entails a three part disjunction, each disjunct of which is implausible. But this does not entail that the disjunction itself is implausible. In fact, I think the friend of determinate identity should insist that the disjunction is quite plausible; its denial in a case of personal fission, she might add, is absurd. Others will disagree, of course. But, unless one has independent reasons for thinking determinate identity implausible—reasons in addition to the implausibility of (1), (2), and (3)—it is hard to see how one could motivate the implausibility of ((1) or (2) or (3)). At any rate, the opponents of determinate identity have yet to do so.

There is a second reason that one may not argue from a case of personal fission, in particular, to the implausibility of determinate identity.[31] Suppose that we grant, merely for the sake of argument, that ((1) or (2) or (3)) is implausible. Would this give us reason to find determinate identity implausible? Only if determinate identity entails ((1) or (2) or (3)). But it is determinate identity combined with personal fission that entails this, not determinate identity alone. What determinate identity alone entails is that if personal fission were to occur, then ((1) or (2) or (3)) would be true.[32]

So determinate identity entails a conditional (this conditional might be thought of as "the friend of determinate identity's interpretation of a case of fission"). If the conditional is implausible, then determinate identity entails the implausible. But is the

conditional implausible? In answering this question, it is not sufficient to assert that its consequent—((1) or (2) or (3))—is implausible. After all, the conditional if I win the million-dollar lottery, I'll be a millionaire is eminently plausible, though it has (trust me) an implausible consequent. It has, in fact, an implausible antecedent as well. And the conditional entailed by determinate identity is like this. For most defenders of determinate identity presumably find it quite implausible that personal fission occurs. All determinate identity commits one to is that if the implausible were to occur (personal fission), then the implausible would follow (((1) or (2) or (3))). But this conditional need not be implausible at all.

So suppose each of the possible outcomes of fission—(1), (2), and (3)—is implausible. What of the inference from this to the claim that determinate identity entails the implausible and is, therefore, itself implausible? This inference fails twice over. For even if each of (1) through (3) were implausible, it would not follow that the disjunction ((1) or (2) or (3)) is implausible. And even if the disjunction ((1) or (2) or (3)) were implausible, it would not follow that determinate identity is committed to anything implausible at all, for determinate identity is not committed to ((1) or (2) or (3)). (Determinate identity is committed, rather, to a conditional which has ((1) or (2) or (3)) as its consequent).

Let us consider one last challenge. Suppose we grant that were personal fission to occur, we would find each of (1), (2) and (3) implausible. So we might add that we would be unjustified and unwarranted, were a case of personal fission to occur, in believing that (1) is true, (2) is true, or (3) is true (this doesn't imply, of course, that we couldn't be justified in believing that the disjunction ((1) or (2) or (3)) is true). And, one might claim, if we could not tell which of these possibilities would be true, then we ought to conclude that there would be no fact of the matter—and thus that determinate identity is false. Put thus bluntly, this smacks of verificationism. But perhaps there is a subtler way to put the point.

How are we to explain our complete inability to judge the pre-fission person's identity over time? Why would any positive judgment that we make seem unwarranted and unjustified? Why does each of (1), (2), and (3) seem so implausible? One possible explanation is that there is no determinate fact of the matter here to judge, or that—as a result of allegiance to determinate identity—we are mistaken about which positive judgments are available.

None of this, of course, has the makings of a deductive argument against determinate personal identity. But the ready explanation of this perplexity available immediately upon rejecting determinate identity might give one a reason to find determinate identity implausible. So the final challenge facing us is to explain why each of (1) through (3) is implausible and why we would be unjustified in endorsing any of them individually were fission to occur, even though, according to determinate identity, one or another of them would be true.

In order to respond to this last challenge, we must first note that certain features of an object inform our judgments about its identity over time. I judge that the car I find in the parking lot this evening is the same car I left there in the morning because it is the same color, it is the same shape, it is located in the same parking space, my key fits the ignition, there is a familiar piece of unopened junk mail on the floor, and so on. Of course, a car's exemplifying these properties in the evening is not broadly logically sufficient for its being identical with the car I parked in the morning. It is possible, however unlikely, that my car was spirited away and an exact replica (right down to the unopened junk mail) was placed in my parking spot.

The features informing my judgments of identity over time fail not only to be broadly logically sufficient for those judgments to be true, they also fail on the side of necessity. It is logically possible that my car be painted, bent out of shape (both its body and the ignition so that my key won't work), cleaned and vacuumed, and moved to another parking lot.

This same point—the fact that our evidence for judgments of identity is not broadly logically necessary or sufficient for that identity—can be seen by reflecting on an obvious truth: we could possibly be led astray by the properties we usually make use of in forming our judgments—not just of automotive identity—but of identity over time of persons and all physical objects.

The above comments show us that there are logically possible cases where we are unable to make judgments of identity over time for familiar objects, cases where our judgments are fully justified but dead wrong, and cases where our judgments come into conflict (the replica could be placed beside my car, and I'd be unable to discern which is really mine).

A case of fission is, I suggest, a situation in which most of the features that could justify or ground our judgments of identity over time deliver conflicting results. It is, in fact, designed to be such a case; a case of fission is one in which, almost ex hypothesi, two persons result that appear, so far as we can tell, to be equally good candidates for identity with the pre-fission person. That such a case is possible is, of course, entailed by the fallibility of our evidence for identity judgments. Fission cases simply confirm (although in a striking and powerful way) what all of us—defenders of determinate identity included—already know: the facts that inform our everyday judgments of identity over time are such that it is broadly logically possible that they lead us astray, or simply fail to lead us. So not only can the friend of determinate identity account for the epistemic puzzlement produced by fission, it is entailed by other claims that she endorses.

I have not provided a positive argument for the claim that determinate identity is true or that it is plausible. Nor do I claim to have defended determinate identity from every objection. However, cases of fission have been widely hailed as showing that determinate identity is, at best, deeply implausible or, at worst, impossible. I have demonstrated that cases of fission show no such thing. There is no good argument that

starts with a case of fission and leads to the conclusion that determinate identity is either implausible or impossible.

NOTES

* Thanks to Andrew Cortens, William Lane Craig, Marian David, Michael Dickson, Anthony Ellis, Thomas P. Flint, A. A. Howsepian, Eugene Mills, Mark C. Murphy, Michael Loux, Alvin Plantinga, Philip Quinn, Michael Rea, Leopold Stubenberg, Peter Vallentyne, Thomas Williams, and Dean Zimmerman for helpful comments.

[1]Parfit's position on this has remained essentially unchanged since Parfit (1971). In Parfit (1984), he argues that questions about facts of personal identity in a case of fission are "empty questions", without true answers. He does say that there is a "best description" of such cases—that neither resulting person is the pre-fission person—but he is clear that this is not meant to describe the determinate facts of the matter (1984, 260). And in Parfit (1995, 15-16), he reaffirms his rejection of the claim that "in every imaginable case, questions about our identity must have answers, which must be either, and quite simply, Yes or No."

[2]I will focus on personal identity, as does most of the literature on fission. But the puzzle here is not so much about personal identity through fission, as it is about the identity through fission of any dividing object (Cf Johnston (1989, 373)). After all, the fission of any object (person, amoeba, starfish) might seem "doubly perplexing" for those who believe that there must be a true answer to questions about its identity in any describable case.

[3]The four-dimensionalist can say that personal fission occurs when two distinct four-dimensional persons share some temporal parts (i.e., share all of their pre-fission temporal parts). Thus understood, fission seems consistent with determinate claims about every fact of identity countenanced by the four-dimensionalist: claims like every four-dimensional person is determinately identical with herself or himself, and determinately

not identical with any other four-dimensional person; likewise for every temporal part or person-stage.

[4]Note that one upshot of Lewis's understanding of fission of four-dimensional persons is that two distinct four-dimensional persons are, before fission, co-located. Is there any reason why the endurantist cannot follow Lewis's lead here? Is there any reason why she should not say that, before fission, two enduring persons are wholly co-located, but at fission go their separate ways? Two responses are in order.

First, this move is much less plausible given an endurance ontology than it is given an ontology of four-dimensional objects. For when the four-dimensionalist says that two persons are co-located before fission, she means that two persons share some (but not all) of the same proper parts—all of their pre-fission temporal parts. (Persons who share pre-fission temporal parts can be seen as a temporal version of Siamese twins.) But if the endurantist claims that, before fission, two persons are wholly co-located, she is committed to the claim that, before fission, two distinct persons share all the same parts. This is less plausible than saying that two persons share only some of the same parts.

And more importantly, I will argue that we can maintain both endurance and determinate identity without relying on anything as metaphysically suspect as complete co-location of distinct persons. Perhaps if the arguments from fission forced the endurantist either to reject determinate identity or to embrace pre-fission co-location of distinct persons, the right move would be to embrace co-location. Happily, we will see that such a choice need not be made. We will see that the fission argument fails, and that this failure requires no desperate measures from the endurantist.

[5]I defend the claim that endurance is best understood as the doctrine that identity over time is simply identity in Merricks (1994).

[6]Let me say a bit more about why the endurantist, if she embraces determinate identity, must endorse determinate identity over time.

Suppose all facts of identity are determinate. So, for any person, whether that person is identical with P is either determinately true or determinately false. Likewise for whether that person is identical with P*. So, assuming determinate identity, for each person existing at t, there is a fact of the matter as to whether that person is identical with P or identical with P*. Likewise for each person existing at t*. Thus it is either determinately true or determinately false that a person existing at t is identical with P and P* and a person existing at t* is identical with P and P*.

How is this relevant? In Merricks (1994) I argue that the claim that P at t is identical with P* at t* is equivalent to the claim that P is identical with P* and exists at t and t*. We just saw that the friend of determinate identity is committed to the determinate truth or determinate falsity of the claim that P is identical with P* and that P/P* exists at t and t*. So we saw that she is committed to the determinate truth or determinate falsity of the claim that P at t is identical with P* at t*.

[7]I will also assume that persons are composite physical objects. The possibility of fission seems to threaten the powerful intuition that there are—and must be—determinate facts of the matter as to one's own identity over time. Thus some philosophers have found in fission a motivation for dualism, since unextended partless souls would be incapable of undergoing fission (cf. Swinburne in Shoemaker and Swinburne 1984); even Parfit agrees that if persons were "Cartesian egos" there would be determinate personal identity (1984, 273). If the arguments of this paper are successful, we can conclude that dividing composite persons can enjoy determinate identity over time, and so fission need not push anyone toward dualism.

[8]It is only fission resulting from double half brain transplant that I will discuss in this paper. The problems raised for <u>identity</u> (as opposed to future-directed concern) by (so-

called) fission resulting from, e.g., teletransporters gone haywire, are, so it seems to me, much less interesting and more easily dealt with.

[9]Mark Johnston (1989, 377) defends this very strong claim, adding that Parfit fails to recognize the full force of the fission argument.

[10]One might try to defend determinate identity by arguing that a case of personal fission of this sort is impossible. David Wiggins (1976) argues for the logical impossibility of personal fission; his argument presupposes the necessity of natural kind membership, the dependence of natural kinds on the laws of nature, and the nomological impossibility of personal fission.

[11]On the possibility that each of the resultants of fission are proper parts composing a single person Parfit says:

> After I have had this operation, the two 'products' each have all of the features of a person. They could live at opposite ends of the Earth. Suppose that they have poor memories, and that their appearance changes in different ways. After many years, they might meet again, and fail even to recognise each other. We might have to claim of such a pair, innocently playing tennis: "What you see out there is a single person, playing tennis with himself. In each half of his mind he mistakenly believes that he is playing tennis with someone else" (1984, 256-257).

[12]Emphasis added. Obviously, being "exactly similar" cannot imply having <u>all</u> the same properties, since, e.g., one resulting person is in a place not occupied by the other.

[13]So 'P is identical with R' means something in the neighborhood of 'P is (identical with) a survivor of fission (or single half brain transplant) and has a brain composed of only the parts which, prior to fission (or single half brain transplant), were on the right side of P's brain'.

[14]One might instead argue that since the fission resultants are "exactly similar", they both have the "identity-making" property or properties. But this would entail the contradiction that two distinct (i.e., non-identical) persons would, in virtue of their identity with the pre-fission person and the transitivity of identity, be identical with each other. This is, in fact, the very sort of argument discussed in the next section.

[15]A third constraint on this collection of features is that satisfying the first two constraints cannot entail the truth of (1) or (3). For example, the defender of the fission argument could not claim that neither fission resultant is identical with the pre-fission person because an object cannot survive the loss of some of its parts. This would have the result (1) is true, and thus no reductio against determinate identity would be available. Similarly, one could not both defend the fission argument, and follow Nozick in rejecting (2) on the grounds that both resultants lack the (necessary and) sufficient condition of being the closest continuer to the pre-fission person. (Nozick, 1981, 34). If Nozick is right, then (1) is determinately true and determinate identity unscathed.

[16]This is not, however, a premise that will seem obviously true to every friend of determinate identity. For instance, the mereological essentialist holds that no object can be such that it has some parts (such as the atoms composing the left half of its brain) at one time and then lacks them at another. Below I address other reasons one might reject (4) and (5).

[17]Cf. Mark Johnston who says that granting that fission is possible just is to grant that:

> It is conceptually possible that distinct processes which in other possible situations secure the survival of an individual co-occur and when they co-occur they are intrinsically exactly as they were in the respective possible situations (1989, 379).

[18]Note that in order for (4) and (5) to be supported by intuitions that P would survive half brain transplant, it must be the case that the nearest world (or worlds) in which R is in C

and the nearest world in which L is in C* are worlds in which half brain transplant, and not fission, occurs. This is nowhere near obviously true, but I will, for the sake of argument, grant this to the defender of the fission argument.

[19](8) is necessarily false because it entails the contradictory claim that, possibly, P is a person who has received only the left half (and not the right half) of P's original pre-fission brain and P is also the person who received only the right half (and not the left half) of P's original pre-fission brain.

(8)'s necessary falsehood does not follow, as one might initially think, from a violation of the necessity of identity, because 'L' and 'R' are definite descriptions. Compare: although the smartest person in town and the mayor are not identical with S, it is possible that they be; that is, it is possible that S be both the smartest person in town and the mayor.

[20]In the service of clarity, I present this argument as a reductio that results in (8). But (7) is also impossible, since (7) is a subjunctive conditional with a necessarily false consequent, but a possibly true antecedent.

[21]If this argument is successful and proves determinate identity leads to contradiction, one might wonder how even the foe of determinate identity can avoid absurdity. By denying that all questions of personal identity have determinate answers, one could hold that (4) and (5) are neither true nor false, but "indeterminate", thus one is not thereby committed to the claim that (8) is definitely true (see van Inwagen, 1988). Also, one might be a "reductionist" and claim that (4) through (6) should ultimately be translated into claims about, e.g., atoms and, when so translated, produce no contradiction. If something along these lines were true, possibly dividing objects like persons might not have determinate identity (although, e.g., atoms could). Reductionism, combined with four-dimensionalism, is the approach to personal fission opted for by Parfit (where the person is reduced not to atoms, but, roughly, psychological states) .

[22]Cf. Garrett (1990).

[23]Johnston (1989) and Noonan (1985).

[24]One might say P does not exist at this point, but ceases to exist at removal and comes back into existence at transplant. One problem with this response is that some (not me) think intermittent existence is impossible. Another (I think more serious) problem is that if we grant that C** results in P's not existing, then I think we lose much of the initial intuitive support for thinking C*—which has C** as its initial segment—is a process that necessarily secures the continued existence of P. Of course the enemy of determinate identity might believe that there is no determinate fact of the matter as to whether P exists at the end of C**; but she can't assume this—that is, assume that determinate identity is false—in order to defend The Process Principle, and then use The Process Principle to attack determinate identity; that would be question-begging.

[25]To state the obvious: merely rejecting the Process Principle does not commit one to anything like a Best Candidate Theory of identity over time.

[26]Of course, if one rejects (4) and (5), then one might be able to remain wedded to the Process Principle (recall that the argument against that principle begins with the assumption that (5) is true). But if (4) and (5) are false, then the fission argument cannot even get started. My central aim here is to block the fission argument, not to demonstrate the falsity of the Process Principle.

[27]One might suggest that (4), (5), (4*), and (5*) are not the culprits here. One might suggest that the puzzle I've been discussing confronts anyone who thinks that an object can lose parts, not just those who endorse (4) and (5). I disagree. For while I find it quite plausible to follow van Inwagen (1981) and deny the existence of "arbitrary undetached parts" like the whole of my body minus a hand, or the whole of a cat minus its tail, (4) and (5) force something more troubling on us. The eliminativist response to this puzzle,

if we embrace (4) and (5), asks us to deny the existence of what we might call a "non-arbitrary" part, a brain hemisphere.

[28]Above I said that pre-fission co-location, given four-dimensionalism, is more plausible than pre-fission co-location given endurance. And I stand by that claim. But let me add that pre-fission co-location given four-dimensionalism, and thus the standard four-dimensionalist response to fission, is still fairly implausible. As noted above, fission of four-dimensional objects implies that distinct objects can share some of their (i.e. all of their pre-fission) temporal parts. This means that if you divide in the future, then one of two unseemly results follows. One of the possibilities is that there are two persons in the place you occupy right now—this is because (in one sense) what it is to have a four-dimensional person in a place right now is to have that person's present time-slice in that place. The only other option is that you currently are nowhere at all right now—this is because (in another sense) a whole four-dimensional person cannot occupy any place at one time.

[29]This is called the "Paradox of the Preface" because it is clearly illustrated by an author who, in a preface, acknowledges that something or other she says in the book to follow is sure to be false, although it goes without saying that each of the claims in the book is there because she believes it to be true.

[30]I do not, and need not, endorse the very strong claim that, for any p, finding p plausible entails that one finds not-p implausible. Rather, I claim only that it is often, perhaps usually, the case that if a rational person finds p plausible, she finds its denial implausible.

[31]What I say below would not be relevant to a case of, for example, amoeba fission.

[32]Determinate identity, if fission is possible, entails that (1) or (2) or (3) is possibly true. The question at issue is not, however, whether it is plausible that (1) or (2) or (3) could be true—that's a question of possibility, and it was answered in the preceding section. That

(1) or (2) or (3) is plausible means something like it is plausible that (1) or (2) or (3) is, was, or will be true.

REFERENCES

Evans, Gareth. (1978) "Can There Be Vague Objects?" Analysis 38, 208.

Garrett, Brian. (1990) "Personal Identity and Extrinsicness," Philosophical Studies, 59, 177-194.

Johnston, Mark. (1989) "Fission and the Facts," James Tomberlin (ed.) Philosophical Perspectives, 3, (Atascadero, CA: Ridgeview).

Lewis, David. (1976) "Survival and Identity," A.O. Rorty (ed.) The Identities of Persons, (Berkeley: University of California Press).

Merricks, Trenton. (1994) "Endurance and Indiscernibility," Journal of Philosophy 91, 165-184.

Noonan, Harold W. (1985) "The only x and y principle," Analysis 45, 79-83.

Nozick, Robert. (1981) Philosophical Explanations (Cambridge: Harvard University Press).

Parfit, Derek. (1971) "Personal Identity," Philosophical Review, 80, 3-27.

Parfit, Derek. (1984) Reasons and Persons, (Oxford: Oxford University Press).

Parfit, Derek. (1995) "The Unimportance of Identity," Henry Harris (ed.) Identity, (Oxford: Oxford University Press).

Shoemaker, Sydney and Swinburne, Richard. (1984) Personal Identity, (Oxford: Basil Blackwell).

Van Inwagen, Peter. (1981) "The Doctrine of Arbitrary Undetached Parts," Pacific Philosophical Quarterly 62, 123-137.

Van Inwagen, Peter. (1988) "How to Reason About Vague Objects," Philosophical Topics 16, 255-284.

Wiggins, David. (1976) "Locke, Butler, and the Stream of Consciousness: and Men as a Natural Kind," A.O. Rorty (ed.) The Identities of Persons, (Berkeley: University of California Press).

Wiggins, David. (1980) Sameness and Substance, (Cambridge: Harvard University Press).

Department of Philosophy
Virginia Commonwealth University
915 West Franklin Street, Box 2025
Richmond, VA 23284-2025
U.S.A.