

Fiction and Fictions: On Ricoeur on the route to the self

Simon Beck

School of Philosophy and Ethics
University of KwaZulu-Natal
Private Bag X01
Scottsville, 3209
South Africa
Email: beck@ukzn.ac.za

Abstract

In reaching his narrative view of the self in *Oneself as Another*, Paul Ricoeur argues that, while literature offers revealing insights into the nature of the self, the sort of fictions involving brain transplants, fission, and so on, that philosophers often take seriously do not (and cannot). My paper is a response to Ricoeur's charge, contending that the arguments Ricoeur rejects are not flawed in the way he suggests, and that his own arguments are sometimes guilty of the very charges he lays at the door of his opponents.

Section 1: Introduction

In supporting his narrative view of the self in *Oneself as Another* (Ricoeur, 1992), Paul Ricoeur argues that, while literature may well offer revealing insights into the nature of the self, the sort of fictions involving brain transplants, fission, and so on, and that philosophers often take seriously do not – and indeed cannot. My concern in this paper is not so much with Ricoeur's favoured theory (although it will come up), but more specifically with his attack on these fictions. While I think the issue is an important one in itself, it is perhaps worth noting here that Ricoeur rather raises the stakes by suggesting that this disagreement is not far from the heart of the disagreement between the hermeneutic and analytic traditions in philosophy.

Criticism of the use of fictions in philosophy is not unusual. But criticism of the method from this particular source is, at least initially, surprising. The standard practice for many who propose theories of the self or personal identity is to claim support for their conclusions from *stories* like that of Locke, in which a prince and a cobbler supposedly swap bodies (Locke, 1947:169). Yet Ricoeur is proposing what he calls a *narrative* theory of personal identity (in other words one that will centrally involve stories) while rejecting the method of appealing to these stories out of hand. The attack may seem even more surprising in the light of the fact that Ricoeur agrees that fiction – in the sense of literary fiction - can tell us much about ourselves, and much that is fundamental. His writings on the self are drenched with references and appeals to literature. Moreover, some of what he believes fiction tells us coincides with the conclusions that the criticised philosophers draw from their fictions.¹ Nevertheless, Ricoeur

¹ See, for instance, his footnote 14 on p151: “I have not yet said my final word on Parfit. Later we shall ask whether a certain convergence of the literary fictions which I assign to selfhood and those of science

insists, thought-experiment stories like Locke's are utterly uninformative with regard to the self. In the light of this, we expect a surprising argument to make his case; my point will be that no such argument is forthcoming.

This is how I will approach the issue. First, I will set out an example of a mainstream use of a fiction concerning the self * one of Derek Parfit's thought-experiments, the one that happens to be a focus of Ricoeur's attack. Second, I will set out an example of the use of literary fiction of which Ricoeur approves as a contrast. With this background in place I will be in a position to explain what it is that Ricoeur sees as right about the one and drastically wrong about the other. I will also be in a position to argue that Ricoeur's attack misses its mark altogether.

Section 2: The wrong sort of fiction

The example of the offending use of fiction that I will describe is Parfit's watershed thought experiment that he called "My Division" (Parfit, 1984:254). Parfit used the story in his attempt to show that there are certain principles built into our common-sense understanding of ourselves which we need to give up.

For the purpose of the thought experiment, Parfit assumes that he is one of a set of identical triplets, and that each of his brain hemispheres is capable of the functions of the other. He denies that there is any deep impossibility in the second assumption by pointing out that certain stroke sufferers are able to regain all their cognitive functions, despite not having the use of one brain hemisphere. He then outlines the following scenario.

My Division. My body is fatally injured, as are the brains of my two brothers. My brain is divided, and each half is successfully transplanted into the body of one of my brothers. Each of the resulting people believes he is me, seems to remember living my life, has my character, and is in every other way psychologically continuous with me. And each has a body that is very like mine. (Parfit, 1984:254)

Fantastic though this may be, Parfit insists that it does not go beyond the bounds of empirical possibility. The most dubious part is the splitting of a person's consciousness into two streams - but he argues that this can already be observed in the case of epileptics who have undergone a commissurotomy. He points out that there are only four possible answers to the question, "What happens to me?" in the light of this scenario:

- (i) Nobody who survives is identical with me; that is, in effect, I have died.
- (ii) I am one of the surviving people.
- (iii) I am the other survivor.
- (iv) I am both surviving people.

The trouble is that all of these answers are problematic. It would be strongly counter-intuitive to say that the original person was dead; that is, that nobody is identical to the original. As Parfit puts it, this would amount to calling a double success a failure (1984:256), since in the case of each survivor it is clear that, if he were the only survivor then he *would be* the original. We cannot go for option (iv), since the logic of identity prevents both candidates from being the original. But to say that one of the candidates retains the identity of the original while the other is merely exactly similar,

fiction (which, in my opinion, concern only sameness) is not reconstituted when one takes into account the ethical implications of narrativity. There is perhaps for us, too, a way of saying that identity is not what matters."

is to make an arbitrary, ungrounded decision. However, those are all the available options.

Parfit's response to this dilemma was that we must give up a principle that seems fundamental. For what we are faced with in the result of My Division is a case in which we have everything that could be fundamental to grounding a claim of personal identity, but in which we cannot apply the term "identity," since we have everything that is fundamental twice over. His conclusion is that it cannot ultimately be *identity* that matters in survival. Personal identity is what we have when we have all that fundamentally matters in survival *and only one instance of that relation*. We are obliged to give up the "natural view" that "what is judged to be important ... is whether ... there will be someone living who will *be me*. On (the natural) view, this is always what is important" (1984:215). The "natural view" has been common sense at least since Descartes' time, but we must give it up. The debate should be about what matters, not identity. It might be easier to grasp what has changed by thinking of it in another way – what we are giving up is the view that there is always a fact of the matter as to whether somebody is or is not you. Parfit holds that this apparently crucial question ("Would that still be me?") can be an empty one.

The argument is an important one in the debate where Parfit's work occurs. Parfit takes the rejection of the principle that identity is what matters as a crucial step in opening the way for a reductionist view of the person: it is (in Parfit's view) one of two principles on which non-reductionism – the view that the self is something over and above physical or psychological facts – depends.

Section 3: The right stuff

Before we look at Ricoeur's complaints about arguments like the one just outlined, let us take a look at a way in which he thinks fiction can enlighten us on the self. Ricoeur points to certain "unsettling cases" in contemporary plays and novels:

To begin with, these cases can be described as fictions of the loss of identity. With Robert Musil, for example, *The Man without Qualities* – or more precisely, without properties ... – becomes ultimately nonidentifiable in a world, it is said, of qualities (or properties) without men. The anchor of the proper noun becomes ridiculous to the point of being superfluous. The nonidentifiable becomes the unnameable. To see more clearly the philosophical issues in this eclipse of the identity of the character, it is important to note that, as the narrative approaches the point of annihilation of the character, the novel also loses its own properly narrative qualities ... To the loss of the identity of the character thus corresponds the loss of configuration of the narrative (Ricoeur, 1992:149).

Ricoeur contends that cases like this show a constitutive relationship between a narrative and a self: "these unsettling cases of narrativity can be reinterpreted as exposing selfhood by taking away the support of sameness" (1992:149).

Section 4: Why thought-experiments might fail where literature works

What is it that is wrong with the story of My Division? Ricoeur dismisses stories like that as "*technological* fictions": they represent a "technological dream" that does not reflect reality in the way that *literary* fictions (like Musil's) do. The crucial aspect of missing reality is *embodiment*. Literary fictions are all consistent with the requirement for embodiment, and take it as just that – a requirement – a necessary feature of our

existence. “Literary fictions differ fundamentally from technological fictions in that that they remain imaginative variations on an invariant, our corporeal condition experienced as the existential mediation between the self and the world” (1992:150). Ricoeur's contention is that this ingredient is missing from all the fictions in question. Authors circumvent the problem by appealing to technology, and not just technology, but “dream technology” – mechanisms that are not available in reality. Thought-experiments like Parfit's end up with reductionist conclusions that suggest that our embodiment is merely contingent through an appeal to dream technology, but in the end this is simply the result of a begged question.

To be fair, we need to take these points more slowly (although they do run together in Ricoeur's own exposition of the problem). Ricoeur's case does not actually seem to turn, as the previous paragraph might suggest, on the point that the imagined scenarios are impossible in the sense that we do not yet have sophisticated enough technology to bring them about. If that was all he was saying, then his opponents would have a fairly easy time responding (as opposed to the harder time they would have explaining away logical or some other deeper kind of impossibility).

Rather, his point appears to be that the willingness to contemplate such scenarios is a symptom of a deeper problem. He argues that thinkers in the Locke/Parfit tradition focus their discussion on the issue of “idem-identity” – that is, on factors relating to sameness of character over time, and not on “ipse-identity” or selfhood. It is in regard to selfhood that the requirement that we be embodied (and own our bodies) emerges, and it is this notion of identity that is important. “The hermeneutics of existence,” he says, “takes corporeality to be insurmountable” (1992:150): he accepts with the tradition that you cannot be an agent or a sufferer (something that things happen to) – both hallmarks of persons – unless you are embodied. Perhaps this does not render his point accurately; the point is better captured by his claim that “the factual character” of certain mental and physical events involves a “phenomenon of mineness” (1992:132). This is the form of non-reductionism that Ricoeur espouses. He objects to Parfit's characterization of non-reductionism as the view that personal identity is a matter of a further fact over and above mental and physical facts, since that already (and illegitimately) classifies mental and physical facts as *impersonal*. In his eyes, then, the whole focus of the arguments by Locke and Parfit is on the wrong notion of identity.

He believes that the misguided focus of Parfit's analytic tradition is reflected in their thought experiments. He complains (1992:150) that they ignore the *person* and take the brain as representative of the person, and then apply their dreams of bisection, transplant, reduplication, and so on, to it. In this way, technological fiction leads them to impersonal, reductionist accounts of identity – but only because that impersonal presumption against embodiment was there in the first place. In effect, they take embodiment to be contingent, and only then is there conceptual space for the thought-experiments to “work”. As a result, while Ricoeur is willing to agree that there are aspects of the self that are contingent despite the counter-dictates of common sense, he cannot accept that its contingencies are to be discovered by the method of thought-experiment (nor that there are any contingencies about embodiment to be discovered).

Section 5: Why thought-experiments are innocent – in principle

The charge of begging the question – even if Ricoeur does not use that term – is a serious one. However, it is much harder to see whether it can seriously be held against the arguments in which the offending thought experiments are embedded. Locke's prince

and cobbler story is part of an answer to the question, “Does our concept of personal identity, as it stands, require sameness of body (or human)?” Locke produces evidence in terms of a case in which we agree that there is personal identity, yet where there is no sameness of body: when a soul (together with a psychology) move from one body to another. Since we understand the case, and agree readily about who is who, we have evidence that the limits of our concept have not been transgressed. In other words, we have evidence that disembodiment (in the sense of not necessarily owning *a particular body*)² is consistent with our understanding of personal identity. Ricoeur charges his opponents with “inversion of meaning”, but Locke is making meaning explicit rather than inverting anything. It is true that the cobbler ends up with the prince's character (there is idem-identity), but there is no reason to accede (at least without much more argument – that is, without many more appropriate thought-experiments or psychological evidence) that this was all there was, despite it clashing with our deep beliefs, which led us to agree that it is now the prince in the cobbler's body.

Unless there is some further argument to show that something in the way the case has been described unfairly misleads us or hides some conceptual impossibility, there are no grounds for the charge that the case *presupposes* disembodiment. We would have to be shown where this presumption has been illicitly sneaked in – otherwise what we have is a case against the requirement for embodiment (understood as the requirement that only this body can have the feature of being *mine*). To attempt to block the argument on the grounds that it is willing to consider disembodiment is to beg the question the other way – to assume (illicitly) that embodiment (as in owning a particular body) *is* a requirement when that is what is up for grabs.

Friends of Ricoeur might wish to object at this point that he does not simply presuppose embodiment as I suggest, but rather that his claims are to be understood in the broader context of his hermeneutical project. In the first place, however, this is not how he presents his case against technological fictions. He addresses himself to thinkers like Parfit and argues, as I have outlined, that the fictions in question are illegitimate, not that they are *illegitimate in the hermeneutical project*. In the second place, if we are required to understand the present claims against the background of his broader project, then this sounds precisely like a presupposition of embodiment. That is, it is a presupposition, unless there are independent reasons for the users of technological fictions to accept Ricoeur's project. Yet Ricoeur appears to be presenting his case against these fictions as a reason for giving up the analytic project to join him in his hermeneutic one.

Before I go on to try and turn the tables on Ricoeur, I need to return to one point I made earlier. I defended Locke against Ricoeur's charge of “inverting meaning” in his thought experiment on the grounds that he was only relying on commonly shared meaning. The same defence cannot be offered when it comes to “My Division”, since Parfit is openly arguing that our common meaning *needs* to be inverted. Nevertheless, I do not think that Parfit begs the question (at least not in the way that Ricoeur suggests he does), and that is what I will argue in the next section.

2 I am aware that Ricoeur means much more than this by embodiment: he believes that it is a fundamental feature of our experience that we experience ourselves as embodied in a particular way (and thus that when we think we are imagining things to be otherwise, we are not really imagining that at all). All the same, his claims about embodiment do imply that each of us necessarily owns a particular body – the claim I am focussing on.

Section 6: Why thought experiments are innocent in detail as well

I argued that thought-experiments need not in principle be guilty of Ricoeur's charge, and I think a defence can be offered of the detailed argument against particular examples as well. He offers as evidence of the "impersonal" assumption by the opposing tradition, the following aspect of their thought experiments.

In this (technological) dream, the brain is taken to be the substitutable equivalent of the person. The brain is the point of application of advanced technology. In experiments of bisection, transplantation, reduplication and teletransportation, the brain represents the human being as the object of manipulations (Ricoeur, 1992:150).

Now, someone like Parfit (as well as someone like me and, I suspect, someone like you) might well think it worth considering whether a brain can be a person; but that aside, it is still hard to see what the illicit assumption is that Ricoeur sees in a reduplication thought experiment. The question is, "What is the relation between the two people that result and the original persons?" There were three people before and two people now – what can we say has happened to those original three in identity terms? Ricoeur charges that the brain is illegitimately taken as representing the person – but there is nothing in the case that fits the description of "the person" that any brain is taken to substitute. The questions are asked of whole embodied people – Parfit and his brothers beforehand, Lefty and Righty (the two surviving and entirely functioning people) after. Brains (or parts of them) have merely been functional in supporting psychological states – and nobody denies that they do that. Parfit's case is set up to embarrass someone who claims (like Ricoeur) that this body and psychology are irreducibly mine, just as much as someone who claims (like Descartes) that the self is something over and above the relevant mental and physical events: the facts of My Division (seen personally or impersonally) cannot happily fit into Ricoeur's categories. We can put Ricoeur's problem in his own terms, like this: after My Division, it is either the case that Derek does not exist, or one of those sets of mental and physical events is (irreducibly) *his*, even though we cannot know which set it is.

That it is not Locke and Parfit who beg the question, but rather Ricoeur who does so, is further suggested by the distinction the latter insists on between real literature and the fiction of thought experiments. Real literature, he contends, can reveal contingencies concerning the self. In the stream-of-consciousness novel, "we ... reach an extreme pole of variation where the character in the story ceases to have a definite character. It is at this pole that we encounter limiting cases in which literary fiction lends itself to a confrontation with the puzzling cases of analytic philosophy" (1992:148-9).

Now we may well want to distinguish between Parfit's *My Division*, or Locke's prince and cobbler story and Robert Musil's *The Man Without Qualities* on all sorts of aesthetic grounds. But on Ricoeur's account, even if Robert Musil were to write *My Division* – the novel – it would not count as literature. It would not count because it would not be an imaginative variation on the required invariant – embodiment. But this reveals the absurdity of the case against thought experiments: they do not reveal that embodiment is contingent because they are not literature, and they are not literature because they do not presuppose embodiment. Literature is defined as that which agrees with Ricoeur's view of the self, and only literature can reveal the true view of the self; unsurprisingly, the true view turns out to be Ricoeur's.

This is enough to defend thought-experimental fiction against Ricoeur's complaints.

But there is still room to say something about the positive spin that Ricoeur gave to literature at the expense of lesser fiction. Some argument is needed to show why thought experiments might support conclusions where Ricoeur said they could not; but it needs very little to cast doubt on whether his literature shows anything at all about the self. The matter is simple: the fact that Musil brings it about that the plot (in *The Man Without Qualities*) vanishes as the person does, tells us at most about his art; one case – or even many unsettling cases – will never be in a position to show that you and I are just narratives, or to show any such thing that Ricoeur believes about the self. Cases can show that a supposedly necessary link is not one, but they will never be in a position to show that a link is indeed necessary in the way that Ricoeur envisages for that between narrative and self. And this holds for the hermeneutics as well as for the analytics.

References

- Locke, John. 1947. *An Essay Concerning Human Understanding*. London: J.M. Dent & Sons.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press.
- Ricoeur, Paul. 1992. *Oneself as Another*, trans. Kathleen Blamey. Chicago: University of Chicago Press.