

# Why Parfit's Contradiction Makes Me Think I Don't Exist

*Alice Evans*

Although counter-intuitive and apparently contradictory consequences may in themselves be insufficient grounds to reject a theory, realising that a theory has such consequences may prompt us to reconsider its premises and question whether they are as sound as we would have otherwise been inclined to believe. If those premises are insufficiently forceful to make us amend our intuitions, so that they accommodate the consequences of that theory, we may justifiably reject that theory altogether. This may be the case for Parfit's theory of personal identity, as revealed by an analysis of his **Division**. I shall argue that because the uniqueness condition is neither logically nor metaphysically justified, Parfit should revise his theory. But such revision entails an inescapable problem. I then provide further reasons to think doubt that Parfit's Reductionism is our concept of identity.

According to Parfit, numerical personal identity is a metaphysical fact that holds because of certain impersonal circumstances. For Parfit, "our identity over time just involves (a) Relation R – psychological connectedness and/ or psychological continuity – with the right kind of cause [ which he takes to be 'any cause'<sup>1</sup>], provided (b) that this relation does not take a 'branching' form, holding between one person and two different people."<sup>2</sup> But discussion may suggest that Parfit be advised to revise his view.

## Parfit's Division

Recall Parfit's example of his own body and his twin's brain being fatally injured.<sup>3</sup> Parfit maintains that he would survive if his brain were transplanted into his twin's body. He also maintains that he would survive upon the transplantation of just one of his cranial hemispheres, if the other hemisphere

---

<sup>1</sup> Parfit, D., 1984, p. 215.

<sup>2</sup> Ibid., p. 216.

<sup>3</sup> Ibid., p. 254-55.

### *Why Parfit's Contradiction Makes Me Think I Don't Exist*

were destroyed. But what if the other hemisphere was not destroyed but instead transplanted into the body of a different identical triplet? Because both offshoots would be psychologically connected with Parfit, both offshoots would not only believe themselves to be Parfit but the requisite relation R would also hold between Parfit and both offshoots. But who are they and what has happened to Parfit?

Consider the possibilities that: (1) Parfit does not survive; (2) he survives as one of them; (3) he survives as the other; and (4) he survives as both. Or perhaps in there are not different possibilities. According to Parfit, it could only be the case that there are these different possibilities if we are separately existing entities, such as Cartesian Egos<sup>4</sup>. But this requires further justification; for, in general, Parfit does not think we must accept Non-Reductionism before we acknowledge that there is a real answer to whether Parfit at  $t_1$  is identical to X at  $t_2$ , because he thinks our answers will represent different metaphysical possibilities. Indeed, as Parfit admits it is only “since [Parfit's Reductionism] recognises other cases where this is a real difference [that] it preserves and explains the truth that these [i.e. identity and non-identity] are different concepts.”<sup>5</sup> For if it is an empty question whether two persons are or are not identical when there is strong psychological connectedness, Parfit would not have distinguished his position from Eliminativism. So given that Parfit's “[Reductionism] does recognise cases where there is a real difference between numerical identity and exact similarity,”<sup>6</sup> he must provide reason why his **Division** is not such a case. Parfit's main example of an empty question is certain Spectrum cases, whereby relation R does not obviously hold and his theory does not thereby appear to commit him to a particular answer. But this same reasoning cannot be used to argue that his **Division** poses an empty question, since we can stipulate that relation R certainly does hold. Although Parfit thinks that “how we choose to describe this case has no rational or moral

---

<sup>4</sup> Parfit, D., 1984., p. 258.

<sup>5</sup> Ibid., p. 272.

<sup>6</sup> Ibid., p. 272.

## *LYCEUM*

significance,”<sup>7</sup> given that a certain answer may be required by Parfit’s theory, answering the question is important, in so far as it betters our understanding of Parfit’s theory.

So recall the aforementioned possibilities. If Parfit could have survived with one hemisphere, had the other been destroyed, it seems implausible that “double success [could] be a failure.”<sup>8</sup> Parfit thinks that (1) is only implausible on the assumption that identity is what matters. But even if identity is not what matters, it still seems rather odd that what would otherwise ensure one metaphysical possibility, namely survival, does not if there is replication. The idea that only one of the offshoots would be Parfit is likewise peculiar, for, in each case, had the other hemisphere been destroyed, Parfit’s theory would declare our protagonist to have survived, in the remaining offshoot.

There is a fourth alternative, that Parfit survives as both. Now such a possibility may “involve a great distortion our concept of a person”<sup>9</sup> but it does seem to be implied by the most plausible version of Parfit’s theory. Because the requisite relation R holds from Parfit to the two resulting offshoots, consistency plausibly requires that Parfit admit that he would be numerically identical with both offshoots.

One might follow Parfit in objecting that “[one] cannot be identical with two different,”<sup>10</sup> but describing them as such, without further argument, begs the question. Neutrality demands that we label them “offshoots” and understand that each has a distinct stream of consciousness. Parfit’s **Division** is thereby analogous to his **Physics Exam**, in which his mind is similarly divided into two streams of consciousness, with each separate state of awareness having different experiences<sup>11</sup>. When discussing his **Physics Exam**, Parfit rejects the

---

<sup>7</sup> Parfit, D., p. 265.

<sup>8</sup> Ibid., p. 256.

<sup>9</sup> Ibid., p. 256.

<sup>10</sup> Ibid., p. 256; this assertion is also found in: Lewis, D., 1983, pp. 61-63; Noonan, H., 1989, p.166; Sider, T., 1996, p.434.

<sup>11</sup> Ibid., 1984, p. 250.

### *Why Parfit's Contradiction Makes Me Think I Don't Exist*

view that a person is the subject of one set of experiences and affirms that “a person could have a divided mind,”<sup>12</sup> with several co-conscious experiences.<sup>13</sup> Parfit thinks that “we can come to believe that a person’s mental history need not be like a canal, with only one channel, but could be more like a river, occasionally having separate streams.”<sup>14</sup> These comments are consistent with (4). Given that both streams are psychologically continuous with the original Parfitian stream, they are both, on the most plausible version of Parfit’s theory, numerically identical to Parfit. That the two streams are housed in the same or different bodies is not plausibly relevant, yet this is the only difference between Parfit’s **Physics Exam** and his **Division**.

#### **The Uniqueness Condition**

Parfit’s reason for thinking that his **Division** offshoots are two different people, neither identical with himself, is that the structure of personal identity “logically” requires that relation R holds uniquely<sup>15</sup>. However, this is not required: identity would not have this logical structure if particular persons were thought of as types. The numerical identical of types is not such that there can only be one of it. Suppose that one horse and two foals are snuggling in a stable. How many types of animal are there? Because the numerical identity of ‘horse’ plausibly takes a type form, we think that there is one type of animal in the stable, multiply realised in three bodies. If each person were likewise a type (namely their own), personal identity would have a logical structure that allows for two people to be numerically identical without being qualitatively identical. Because there is a view of persons that allows for persons to be multiply realised in different humans, it cannot be claimed that the non-branching condition is forced upon us by the logical structure of identity.

---

<sup>12</sup> Parfit, D., 1984, p. 256.

<sup>13</sup> Ibid., p. 250.

<sup>14</sup> Ibid., p. 247.

<sup>15</sup> Ibid., p. 267.

## *LYCEUM*

Since the non-branching condition is not logically required, Parfit must appeal to metaphysical reasons. If it is assumed, as it is by Parfit, that two consciousnesses can be numerically identical when they occupy different temporal locations then the plausibility of the uniqueness condition requires that spatial location is relevantly disanalogous to temporal location. But nowhere does Parfit argue for the requisite disanology between spatial and temporal location and sensibly so, for this view is neither plausible nor consistent with Parfit's arguments on other occasions. For example when Parfit discusses what matters he maintains that "the only difference in the case of division is that the extra years are run concurrently... Double survival is not the same as ordinary survival. But this does not make it death."<sup>16</sup> Indeed, it is because Parfit denies this disanology that he is able to affirm that he is identical with both streams of consciousness in his aforementioned **Physics Exam**. To insist on the uniqueness condition when the streams occupy different bodies seems unjustifiably inconsistent.

Uniqueness does not seem to be metaphysically relevant, as affirmed by Parfit in his discussion of the **Branch Line Case**; for apparently "it makes little [metaphysical] difference that [his] life briefly overlaps with that of [his] Replica."<sup>17</sup> Notably, Parfit does suggest that if that the overlap were more extensive then said identity relations would be questionable. But this is implausible. If the two offshoots are numerically identical in the **Branch Line Case** for ten minutes, adding another two hours, or a week, or a month, cannot plausibly affect their respective metaphysical properties. So, just as Parfit explicitly rejects that "the presence or absence of [a relation holding uniquely] make[s] a great difference to the value of relation R,"<sup>18</sup> I deny that uniqueness affects the offshoots' metaphysical properties.

A reason for denying the metaphysical relevance of uniqueness lies in the intuitive claim that identity cannot be extrinsically determined. Now, on

---

<sup>16</sup> Parfit, D., 1984, p. 262.

<sup>17</sup> Ibid., p. 289.

<sup>18</sup> Ibid., p. 263.

### *Why Parfit's Contradiction Makes Me Think I Don't Exist*

Parfit's account, his synchronic identity is determined by his psychology, an intrinsic property. Because the duplication of one person's synchronic identity conceptually generates diachronic identity, it seems plausible to infer that the identity relations between Parfit and his offshoot Y are determined by their intrinsic properties, namely in virtue of their sharing the requisite essence. Factors extrinsic to this relationship, such as the existence of offshoot Z, do not seem relevant to the issue at hand, namely the adjudication of whether Y shares Parfit's essence. The uniqueness clause seems to be over and above what is required, which seems inconsistent with Parfitian Reductionism, which purports to be able to inform you of the identities of two persons from physical and psychological facts about them. Of course, extrinsic factors can be relevant, in some situations, such as when we try to determine whether A or B is most *qualitatively similar* to Parfit, by some particular axiom; such as when the Warden of All Souls College seeks Parfit's replacement. Because only one person could be *most* qualitatively similar, factors extrinsic to each candidate are relevant, for this question of maximum qualitative similarity calls for *relative* assessment. But because many consciousnesses could be numerically identical to Parfit, such as over the course of time, extrinsic factors are thereby irrelevant to questions of numerical identity.

#### **Counter-Intuitive Consequences?**

Of course the stipulation that numerical identity only holds where there is no branching enables Parfit to avoid the counter-intuitive consequence that a pre-division person is identical to both offshoots. And perhaps this consequence is reason to grant that uniqueness makes a metaphysical difference and thereby deny that numerical identity obtains in branching cases. Parfit certainly seems to appeal to this reason when he denies that he is identical to both offshoots. He apparently has difficulty in contemplating the idea that two offshoots might fight in a duel, for, in the event of death, it would apparently be unclear whether

## *LYCEUM*

there had been a murder or a suicide, or both<sup>19</sup>. But Parfit is too hasty; the consequences of both duelers being numerically identical are not so totally outrageous to justify the ad hoc addition of the uniqueness condition. In fact, such an occurrence is consistent with Parfit's described **Physics Exam**. If, during Parfit **Physics Exam** one of his streams of consciousness realised that it was totally under-prepared, promptly became suicidal and killed the body it resided in, there would be a death. It may be unclear how exactly we should describe this event, for it could be plausibly described as both a suicide and a murder, but it is some form of death all the same (and similarly so for the two offshoots, in Parfit's **Division**). It thus seems that Parfit's uniqueness clause is ungrounded; it is neither logically required nor has metaphysical relevance, but is simply an ad hoc addition to avoid counter-intuitiveness. It thus seems that Parfit should abandon his uniqueness clause, revise his account of identity (to be relabeled 'Parfitianism<sub>2</sub>') and accept (4).

### **Counter-Intuitive and Contradictory Consequences**

The above discussion suggests that the most plausible Parfitian position is that Parfit survives as both offshoots (Y and Z). Thus far this consequence is not that implausible but it becomes increasingly problematic when we consider the offshoots' futures. Suppose that Y becomes a troll and Z becomes a pixie, by some sufficiently gradual process to ensure diachronic identity. Now because there is psychological continuity, the troll is numerically identical with Y and the pixie is numerically identical with Z. But if the troll is numerically identical to Y, who is numerically identical to Z, who is numerically identical to the pixie, there is a problem. The pixie and the troll are neither psychologically connected nor continuous with each other. So, lacking relation R, they cannot be Parfit-ly numerically identical. However, each offshoot is R-related to Parfit and therefore identical to Parfit. Since the pixie is identical to Parfit and Parfit is identical to the troll, the pixie must be identical to the troll. But we have just

---

<sup>19</sup> Parfit, D., 1984, p. 257.

### *Why Parfit's Contradiction Makes Me Think I Don't Exist*

denied that this identity relation is possible, since identity requires relation R. Thus due to the transitivity of identity, Parfitianism<sub>2</sub> commits adherents to upholding this contradiction; that the troll and the pixie are and are not numerically identical

#### **Possible Solutions**

One might attempt to solve this contradiction by denying that the relation of identity is transitive, but this is not plausible; if Parfit is numerically identical to Y who is numerically identical to the troll, there appears to be no way of denying that Parfit is numerically identical to the troll. Another response would be to deny that psychological *continuity* suffices for identity, which could then yield the conclusion that Parfit is only identical to Y, because of connectedness, but not the troll. But given that identity relations are, most plausibly, transitive, such an admission would require the denial that psychological connectedness makes for numerical identity, for if each link in the chain, between Parfit and the troll, is psychologically connected, all are thereby identical. Since Parfit upholds a single occupancy psychological theory of identity, it would then seem difficult for him to deny that psychological *continuity* suffices for identity.

Parfitians could avoid the contradictory consequences by endorsing Lewis's theory that two persons co-exist in the pre-division body, becoming spatially distinct upon division<sup>20</sup>. Lewis defines a continuant person as "a *maximal* R-interrelated aggregate of person-stages,"<sup>21</sup> "each of which is R-related to all the rest (and to itself), and it is a proper part of no other such aggregate."<sup>22</sup> On this basis, Lewis avoids the conclusion, of Parfitianism<sub>2</sub>, that any R-related person stages are I-related [meaning, stages of the same person]. The explicit purpose of Lewis's account is to reconcile the common sense view, that being I-related is what matters in survival, with Parfit's revisionist view,

---

<sup>20</sup> Lewis, D., 1983, p. 63.

<sup>21</sup> Lewis, D., 1983, p. 60.

<sup>22</sup> Ibid., p. 60.

## *LYCEUM*

that being R-related is what matters in survival. But this attempted reconciliation has since been discredited by Parfit, who shows that these two views about what matters are incompatible in fission cases; for a post-fission offshoot may be R-related (and thereby has what matters) to a pre-fission person without being I-related to that pre-fission person.<sup>23</sup> Given that Lewis's theory does not achieve its aim and has many more counter-intuitive implications, Parfitians are unlikely to think it offers an attractive alternative. As such, it appears that Parfitians are inescapably held to a theory that allows for the contradicting troll-pixie situation.

### **Questioning Parfitian Premises**

One might now call into question the Parfitian premises that generate the contradiction. By providing a set of conditions that are largely extensionally equivalent with our common ideas about identity, Parfit's account is relatively plausible. But there are three problems with his argument. Firstly, Parfit's account implies the aforementioned contradiction, which gives reason to doubt *his* account of 'identity'.

Secondly, for there to be people, an account of identity must be both conceptually adequate, i.e. it must talk about what we mean by identity and not something irrelevant, like apples, and it must be substantively true, there must be members of the set it describes. Herein lies another problem with Parfit's concept of identity; it fails to approximate our own. If Parfitian persons are not what we mean to refer to when we talk of persons then Parfit's theory fails for being irrelevant. Consider the following analogy, if someone argued that there were no moral facts, we could not legitimately dismiss this claim merely because it is counter-intuitive. However, if someone argued that there were moral facts and that these were apples, we could reject their account as

---

<sup>23</sup> Parfit, D., "Lewis, Perry, and What Matters" in Rorty, A. (ed.), *The Identities of Persons*., University of California Press: Berkeley, 1976, p. 94.

### *Why Parfit's Contradiction Makes Me Think I Don't Exist*

irrelevant. My contention is that Parfit's concept of identity likewise misses the mark.

A plausible test for whether a theory of identity is conceptually adequate is whether its answer to what matters in survival matches our own answer. Parfit thinks that what matters in survival is that future persons are R-related to us, that they are psychologically connected and or psychologically continuous with us.<sup>24</sup> But Parfit's view is not convincing. Suppose that you are invited to participate in an experiment, in which you will be cloned and one resembling person will die shortly there after. If Parfit is right, that both have what matters to you in survival, then if one dies it should not matter, for the other will survive and he will have what matters for you. But I simply cannot believe that anyone would volunteer for this experiment without the assurance that it would be them who survives. This suggests that identity matters, not relation R.

Parfit objects that there is no good evidence for Non-Reductionism<sup>25</sup> but this is to confuse substantive and conceptual issues. That there are no members for there being any members of the set described by Non-Reductionism suggests that Non-Reductionism is not substantively true (which I grant) but it does not show that Non-Reductionism is not *conceptually* true, that this is not what we mean by identity.

Parfit could reply that although Non-Reductionism may be the correct conceptual claim, because Non-Reductionism is not substantively true we must modify our concept so that it matches what actually obtains. The question then is whether Parfit's view is tolerably revisionist. Parfit may be right that knowing that my clone endures is better than knowing that both of us will die, for at least the survival of my clone allows for the completion of my non-personal projects, such as for global veganism. But the survival of my clone is definitely not as good as my not dying. Finding out that I will die but some clone will survive may be better than nothing but it is definitely not as reassuring as finding out

---

<sup>24</sup> Parfit, D., 1984, p. 217.

<sup>25</sup> Ibid. p.275.

## *LYCEUM*

that I will not die at all. Clearly identity is what matters. Thus Parfit's Reductionism is not conceptually adequate.

Parfit fails to consider that our concept is Non-Reductionist but since there are no such entities, there are no persons. Parfit's failure to consider intension makes him liable to the following error: suppose, for example, that the intension of concept X is "red" and we think that concept has the extension of circles. But further suppose that red does not actually exist. In trying to give an account of X we might successfully give an account which picks out circles, and is thus extensionally equivalent to X. However, it is more accurate to deny that red exists. Although Parfit's Reductionism may provide an account with an extension that is somewhat similar to what we believe the extension of identity to be, this does not show that Parfit's Reductionism is our concept of identity.

I have shown that Parfit should revise his account of personal identity and abandon his uniqueness condition, because it is neither logically required nor metaphysically justified, given his other views. But on this revised account, certain circumstances give rise to a contradiction, in that two streams of consciousness are and are not numerically identical. In my example, the troll and the pixie are each R-related to Parfit and therefore identical to Parfit. Since the pixie is identical to Parfit and Parfit is identical to the troll, the pixie must be identical to the troll. Yet the troll and the pixie are neither psychologically connected nor continuous with each other and lacking relation R they cannot be Parfit-ly numerically identical. I suggest that those who uphold Parfit's revised account, having abandoned the uniqueness condition, are bound to uphold this contradiction; that the troll and the pixie are and are not numerically identical. Parfit cannot dismiss this particular case as representing an empty question for he would otherwise not have distinguished his own position from Eliminativism. Since the alternative, Lewis's cohabitation theory is unable to achieve its own aims, Parfitian identity is in dire straits. Parfit's account is also conceptually unsatisfactory, for it cannot yield the right answer to what matters. Hence Parfit's account is conceptually unsatisfactory, with implausible consequences.

## *Why Parfit's Contradiction Makes Me Think I Don't Exist*

University of Nottingham  
Nottinghamshire, United Kingdom

### **Bibliography**

- Lewis, D. "Survival and Identity" in *Philosophical Papers*, Volume I. New York: Oxford University Press, 1983.
- Noonan, H. *Personal Identity*. London: Routledge, 1989.
- Parfit, D. *Reasons and Persons*. Oxford: Clarendon Press, 1984.
- Parfit, D. "Lewis, Perry, and What Matters" in Rorty, A. (ed.), *The Identities of Persons*. University of California Press: Berkeley, 1976, pp. 91-107.
- Sider, T. "All the World's a Stage" in *Australasian Journal of Philosophy*, No. 74, 1996, pp. 433-53.