

THE IRRELEVANCE/INCOHERENCE OF NON-REDUCTIONISM ABOUT PERSONAL IDENTITY

David W. Shoemaker
California State University, Northridge
Department of Philosophy
18111 Nordhoff St.
Northridge, CA 91330-8253
(818) 677-7501
FAX: (818) 677-5687
e-mail: david.shoemaker@csun.edu

ABSTRACT FOR
"THE IRRELEVANCE/INCOHERENCE OF
NON-REDUCTIONISM ABOUT PERSONAL IDENTITY"

Before being able to answer key practical questions dependent on a criterion of personal identity (e.g., am I justified in anticipating surviving the death of my body?), we must first determine which *general* approach to the issue of personal identity is more plausible, reductionism or non-reductionism. While reductionism has become the more dominant approach amongst philosophical theorists over the past thirty years, non-reductionism remains an approach that, for all these theorists have shown, could very well still be true. My aim in this paper is to show that non-reductionism is actually either irrelevant – with respect to the practical questions we want answered – or logically impossible. In arguing for this conclusion, I draw from a case Derek Parfit has employed – the Combined Spectrum – and I provide a number of variations to it which ultimately reveal that we have no possible rational recourse other than to become reductionists.

THE IRRELEVANCE/INCOHERENCE OF NON-REDUCTIONISM ABOUT PERSONAL IDENTITY

Introduction

What is it that we want from a theory of personal identity? That is, what questions are we looking to have answered by this sort of metaphysical theory? I'm sure there is (and always will be) a small minority of people interested in the problem solely because it is a problem, yet another fun little metaphysical puzzle to chew on in the ivory laboratory for a while. But for the majority of people interested in the problem, I suspect, there is something more significant at stake, something truly life or death about the matter. For them, we *need* a plausible account of identity, not simply for its own sake, but for its service in answering other, crucial questions about our lives. A viable theory of personal identity, then, will be like money: an instrumental good. And while there will always be some philosophical misers on this issue (for whom a theory of identity is solely intrinsically good), most of us want some metaphysical purchasing power out of it.

Specifically, we want a theory of identity to help answer two general questions, one forward-looking, one backward-looking: (a) what, if anything, justifies and/or explains my special anticipation/concern for certain events in the future; and (b) what, if anything, justifies and/or explains the way we treat people now with respect to certain events in the past? And each of these general questions involves more specific sub-questions. For example, (b) includes questions about moral responsibility (what justifies our holding someone morally responsible for some past action?) and distributive justice (what justifies certain distributions of benefits/burdens based on fairness and/or desert?). And (a) includes the biggie: is it possible for me to survive the death of my body (in other words (maybe), to anticipate certain events in heaven, say)? Indeed Joseph Butler remarks that "[w]hether we are to live in a future state . . . is the most important question which can possibly be asked"¹

The relation between these questions and the issue of personal identity seems rather obvious. For person P justifiably to be held morally responsible for action A (at some past time), it must be the case (it seems) that P is identical to the person who performed A, i.e., that P just *is* the A-er.² Furthermore, to be justified in distributing some benefit to P now as compensation for some past burden, it must be the case (it seems) that P is identical to the person who underwent that past burden. Finally, for me rationally to anticipate the possibility of experiencing heaven, it must be the case (it seems) that the person undergoing experiences in heaven would be *me*.

What we want, then, are the conditions under which I now am identical to some future or past person (or person-slice, or person-stage, etc.), so that we can see whether or not certain attributions of moral responsibility, certain distributions of benefits/burdens, and/or certain anticipations of future events are *justified*. If it turns out, for example, that the only viable theory of identity requires persistence of one's body, then such a theory would imply that I cannot survive its death, and so my anticipation of an afterlife is utterly unjustified. And if it turns out that there are no persisting persons, then distributions depending on compensation and attributions of moral responsibility are themselves unjustified. In any case, though, the framing of the questions, and the answers we require, suggest a specific methodology: figure out the proper metaphysical criterion of identity and then see what it entails about the things that matter to us. If we must consequently give up certain commitments as unjustified, then so be it.³

The dominant criterion of personal identity over the past thirty years or so – *reductionism* – is actually a kind of meta-criterion. Instead of offering a specific set of conditions for what it is that makes X and Y (both persons) identical with one another, reductionism in its most general form simply involves a thesis about the sorts of *facts* to which any such criterion should or should not make reference. According to Derek Parfit, the view's most famous exponent, reductionism is the view that the facts of persons and personal identity over time simply consist "in the holding of certain more particular facts" (p. 210) about brains, bodies and interrelated physical and mental events.⁴ If one believes the facts of persons and

personal identity involve some *further* fact(s) (such as a Cartesian Ego or a soul), one is a non-reductionist.

Obviously, though, there are numerous possible criteria of personal identity compatible with the general reductionist thesis. Consider first some variations on a *Physical Criterion*: X and Y are identical if and only if X is physically continuous with Y, where this might mean (a) X's body is continuous with Y's body (however defined); or (b) X's head (and everything it contains) is continuous with Y's head (and everything it contains); or (c) X's entire brain (including the stem) is continuous with Y's brain; or (d) enough of X's brain to sustain life is continuous with Y's brain; or (e) X's cerebrum is continuous with Y's cerebrum. And there may be others. Similarly for the *Psychological Criterion*: X and Y are identical if and only if X is psychologically continuous with Y, where this might mean (a) the cause of this continuity is *normal*; or (b) the cause of this continuity is *reliable*; or (c) the cause of this continuity is *anything*.⁵

Some of these criteria may be more or less equivalent, of course. Psychological continuity with its normal cause, say, once "normal" is specified, may just coincide with one of the last three versions of the Physical Criterion, and the same could go for psychological continuity with a reliable cause.⁶ But the initial point is that reductionism alone does not get us very far with respect to the generation of a *specific* criterion of identity. In addition, an advocate of reductionism *simpliciter* will have no ready answers for the questions that motivated our project. Consider the issue of immortality, for instance. On Physical Criterion (a), I could not survive the death of my body, insofar as it seems impossible for there to be a body in an afterlife that is continuous with my decaying body on earth (and the issue is made even more clear if my body is cremated after I die). But on Psychological Criterion (c), I *could* conceivably survive the death of my body, just in case God, say, constructs a person in Heaven with whom I am psychologically continuous (and God wouldn't even have to be able to do so *reliably*, on this version; even if a bungling God were occasionally to get it right, that would be enough to establish the possibility). Much more work, then, would be needed to specify which of the

several variations of reductionism is correct before we could get any real answers to our motivating questions.

Providing this specification, however, is not my concern here, for before we engage in any such work we need first to find out why reductionism in general is a more compelling metaphysical view than non-reductionism. If non-reductionism could still be true, then a negative answer yielded by the proper reductionist specification of identity regarding the prospects of surviving the death of my body would remain haunted by the specter that such survival is nevertheless possible. For all the reductionist may say that casts doubt on our prospects for immortality, for instance, the non-reductionist may always reply that it is nevertheless *possible* that there is some further fact true of us that could provide the necessary mechanism for getting us beyond this mortal coil. My aim here is to cut off this reply once and for all by showing that the non-reductionist stance is either utterly irrelevant or utterly incoherent. I begin with a few well-known arguments against non-reductionism, which as they stand do not do the trick. I will then offer several variations on a Parfitian thought experiment that collectively do.

Ego-ism

The reason we should be reductionists, Parfit maintains, is that non-reductionism is extremely hard to believe. The non-reductionist, of course, holds that our identity in its nature involves a deep, further fact and does not involve simply the holding of certain more particular facts. So the non-reductionist believes that even when we have gathered all the facts together regarding the body, brain, and experiences of the person in question, we still do not have the key further fact necessary to determine questions of identity. For the non-reductionist, persons involve something more than the sum of their material parts and thus cannot simply be reduced to them. As a further result, questions of identity are, for the non-reductionist, always determinate: identity is not a matter of degree but is, rather, all-or-nothing. The paradigm example of the deep further fact of personhood for the non-reductionist is the Cartesian Ego, a separately existing entity.⁷

Parfit's line against this view is not, as some have maintained, that it is unintelligible, but rather that *it might have been true*. Those who hold that we are separately existing entities claim that the carrier of psychological continuity is the soul, or something like the Cartesian Ego. Parfit feels that evidence could have been given to support such a theory.⁸ For example, suppose a Japanese woman claims to remember having lived a life as a Celtic warrior in the Bronze Age. Based on her apparent memories, many of her predictions could be checked out by archaeologists. She may remember burying a particular style of headdress or bracelet in a certain place, say, and archaeologists might find such an item in an area shown to have been undisturbed for 2,000 years. And this woman may make many other predictions that are also verified.

Then suppose that many other people have similar experiences, and their predictions are also similarly verified. If we had no other way of explaining such phenomena, then, it seems, we would be forced to accept some sort of psychological continuity between, for example, the Celtic warrior and the Japanese woman. And, further, if we were able to verify that there was no physical continuity between the two, then we would have to abandon the theory that the brain is the carrier of memory. We may then have to conclude that the carrier of psychological continuity *is* something non-physical, something immaterial, something very much like the soul, or, more particularly, the Cartesian Ego. This would mean that we are actually separately existing entities: personal identity would not just consist in facts about brains and bodies, but would also have to involve the further fact that we are essentially purely mental entities that can exist independently of bodies and brains. And it would also mean that questions of our identity would always have determinate answers, for these entities would have an all-or-nothing existence.

But of course we have no evidence of such cases. And in fact we have a great deal of evidence that indicates that the carrier of psychological continuity *is* the brain. So while the non-reductionist theory -- the theory that we, as separately existing entities, are essentially Cartesian Egos -- *may have been true*, given certain evidence, we do not have any of the required evidence,

and the evidence we do have seems strongly to indicate that the theory is *not* true. Thus, reductionism is by far the more compelling view.⁹

How convincing is this argument, though? Appealing to evidentiary considerations is a rather tricky matter when it comes to dealing with immaterial substances, and the methodology Parfit employs here might very well beg the question. After all, if the Ego we're investigating is truly immaterial – a purely thinking *thing* – then it is by definition impossible to verify via empirical means. Unextended substances are beyond our empirical ken. For instance, a believer in the Cartesian view might reply to the Japanese Woman argument that souls simply aren't reincarnated – they get assigned one body on earth and are, after that body's death, "transported" to the afterlife. But while on earth they are the true carriers of psychology and are thus the preservers of identity. Looking for psychological continuity between some existing person now and a person dead 2000 years is, therefore, going to be an utter waste of time.

Indeed, rejoinders are available to the Cartesian theorist for any number of so-called evidentiary considerations against the view. Consider, for instance, the claim that the mental is surely dependent on the physical, given the overwhelming evidence we have that psychological changes accompany physical changes to the brain, e.g., drug hallucinations, Phineas Gage-type cases (in which the destruction of a part of the brain produces great personality changes), chemical changes in the brain associated with learning, etc. The true believer is undaunted by such a challenge, however, for it may simply be the case that while psychological changes *accompany* such physical changes, this constitutes no direct evidence that the mental is *dependent* on the physical. When attached to a body, it might be said, a soul's psychological states work in a kind of *harmony* with a brain's physical states, such that changes to the mental occur *simultaneous with* changes to the physical, without there being a causal dependence of the mental on the physical. The reason for such may simply be beyond our understanding, although one suggestion might be that it better enables us to predict each other's actions and reactions in a way that facilitates coordination and cooperation.

At any rate, though, the soul is a slippery little sucker, and its allegedly immaterial nature is precisely what renders evidential considerations against it moot. A different strategy, then, might be to turn the tables on the view, using its immaterial nature against it with respect to personal identity. Specifically, we might try the tack taken by the world's most famous fictional philosopher, Gretchen Weirob, in the First Night of John Perry's A Dialogue on Personal Identity and Immortality.¹⁰ She runs the following *reductio* on what her dialogue partner, Sam Miller, calls the Soul Criterion of personal identity:

1. If the Soul Criterion were true, we could never have the grounds to reidentify people.
2. We do occasionally have the grounds to reidentify people.

3. Thus, the Soul Criterion is false.

In other words, if the facts of personal identity involved further facts about souls, then the judgments of identity and reidentification we make every day would themselves be judgments about souls. If that were the case, though, they'd be utterly groundless, given that souls are by definition unobservable, immaterial substances. But we make such judgments all the time, and they are clearly *not* groundless and unjustified, "so we must not be judging of immaterial souls after all."¹¹ Consequently, any such non-reductionist, further fact view of personal identity must be false.

Unfortunately, while a valiant effort, this argument conflates two senses of "criterion." On the one hand, a criterion of identity (in the metaphysical sense) might be an account of what *makes* X and Y identical. On the other hand, a criterion of identity (in the epistemological sense) might be an account of *how we can justifiably judge* that X and Y are identical. One might try and draw a conceptual link between the two senses of criterion (as Weirob attempts to do), but there is by no means a necessary connection between them. There may, after all, be reliable indirect methods to reidentify people's souls (something Miller tries to establish in the remainder of the First Night of the dialogue, first by positing a 1-1 correlation between bodies and souls,

and then by positing a 1-1 correlation between minds and souls), but there also may *not* be any such methods. Even if there are not, though, this fact does nothing in itself to undermine what might very well be the correct criterion of identity in the metaphysical sense. Let's face it: we may never actually be able to determine whether or not the correct persistence conditions are met in any individual person. Weirob's *reductio* as it stands, therefore, fails to undermine non-reductionism.

Nevertheless, there remains some pull to the argument. For it does seem that if it's truly the case that an account of identity is divorced from our practical epistemological concerns about reidentification, then it is just *irrelevant*. And the Soul Criterion is just that. We reidentify people via their physical and/or psychological features – we lack the ability to do anything else. So even if the Soul Criterion were true, it would be utterly useless for the practical matters of assessing moral responsibility, determining goods distributions, or accounting for future-oriented prudential anticipation. Consequently, for our present purposes, this version of non-reductionism is a dead end, and from here on out, I will simply assume that the only *practically* relevant general view of identity is one that focuses on material relations between bodies, brains, and/or brain-based (mental) events.

Nevertheless, one might maintain a more subtle non-reductionist view, one that holds that yes, we are essentially material creatures, but the facts about our identity do *not* just consist in facts about brains, bodies, etc. Indeed, they consist in some further fact about our *separate existence* from mere brains, bodies, etc. Perhaps we – persons – are something emergent from our material constituents. Or perhaps we occupy a certain privileged perspective as ongoing subjects of experience, entities that *have* brains, bodies, experiences, etc. Citing only the relations between the materials from which we *emerge*, or the materials that we *own* or *have*, therefore, would overlook certain crucial facts about the conditions under which we – separately existing persons – actually persist.

This form of non-reductionism maintains that the "I" involves something more than, or at the very least something different from, the sum of its material parts, even though the "I" itself is

not immaterial. It exists separately from brains, bodies, etc., even though it does not exist *independently* of them (as the soul would, allegedly). It involves a further fact. Now on this version of non-reductionism, the further fact either obtains or it does not. That is, questions of identity always have determinate answers: either "I" exist or "I" do not. The reason for this entailment is that, according to these views, our identity involves *one more thing* than mere particular (numerous, perhaps) facts about brains, bodies, etc., viz., it involves *a fact* about *the* entity which is me, some one thing that is distinct from the mere agglomeration of my material constitution. So given that this further fact is *a fact*, it either obtains or it doesn't. If we are separately existing entities, then questions of identity will always have determinate answers.

The ideal way, then, to refute such a position would be to argue that personal identity is *not* always determinate. If this more subtle non-reductionist (materialist) view is the only remaining non-reductionist view, and if we could show that questions of identity can be indeterminate, then we could safely set aside non-reductionism from this point forward as either irrelevant (in the "soul" version) or straightforwardly false (in the "materialist" version). So let's give it a shot.

Before I Was Garbo: Indeterminacy and the Combined Spectrum

Consider the case Parfit calls the "Combined Spectrum."¹² Suppose I am seated in a special chair where I am hooked up to the machine of an evil scientist, and attached to the machine are, say, 100 switches. If the scientist were to flip one switch, then one percent of my body and brain cells would be destroyed and replaced with one percent of the corresponding cloned body and brain cells of Greta Garbo at the age of thirty. Indeed, let us further suppose (and this develops Parfit's original case slightly) that the scientist has the full blown Garbo Replica (cryogenically preserved, shall we say) seated next to me, and he simply draws from its cells as needed. So, for example, in the 1% case, some of the cells in a specific area of my left elbow might be replaced with cells from that area of the Garbo Replica's left elbow, and some of the cells of my brain might be replaced with cells from a similar area of the Replica's brain. In this case, the person who woke up would perhaps have one new memory and a body slightly

different from mine. If the scientist were to flip two switches, then two percent of my body and brain cells would be destroyed and replaced with two percent of the Replica's corresponding body and brain cells. And so on. If the scientist were to flip all the switches, my entire body would be destroyed and replaced with the Garbo Replica. What would exist in my original chair in this case would be a person with 100% of Garbo's cloned cells, a person henceforth called GarboR.

This case is intended to capture the entire range of possible psychological and physical continuities that may obtain between two person-stages.¹³ The question to consider, then, is who is the survivor at each stage of the spectrum? There are only three possible replies: (a) the survivor would always be me (even in the case where the scientist flips all 100 switches); (b) there is a sharp borderline at some point on the spectrum, before which the survivor would be me and after which the survivor would not be me; and (c) the identity of the survivor would occasionally be indeterminate. This last is, of course, the reductionist response.

One could maintain (a) only if one held that personal identity is entirely independent of any psychological and/or physical characteristics or relations, a position Parfit calls the "Featureless Cartesian View,"¹⁴ and this option is rather easily dismissed as unintelligible. If what makes me me is utterly divorced from any of my psychological or physical features, then it is just as likely that I have a river of qualitatively identical egos/souls running through me than that I somehow survive from day to day in my present body. For practical reasons cited earlier with respect to the Soul Criterion, then, given the utter independence between this view and our desired ability to be able to reidentify people, it is not one able to answer any of our questions with respect to identity. And there are several other absurdities implied by the view which render it unworthy of attention here.¹⁵

But what of (b)? This option allows for what certainly seems obvious: the person at the near end of the spectrum is me, and the person at the far end of the spectrum is not me, i.e., it is GarboR. But the respondent here must claim that there is a borderline at some point on the spectrum, before which the survivor is me and after which the survivor is GarboR (or, perhaps, is

yet a third, different person, a position I will discuss in detail below). But where could such a borderline be? Is it at the 50% mark? At the 51% mark? At the 49% mark? Somewhere else? Even if there could never be any evidence for where this borderline is, anyone holding that the identity of the survivor must always be determinate (whether it's me or not me) must hold that the borderline *is* there, somewhere.

Parfit's reaction to this option is that it is "hard to believe,"¹⁶ for two reasons. First, it is hard to believe that the difference between my life and my death could actually consist in such small differences in my physical and psychological features (one percent, in the case as I have described it). Most of us, after all, believe that the difference between a future person's being me and that person's being someone else is "a *deep* difference."¹⁷ Second, he claims it is hard to believe that there must be a sharp borderline for which we could never have any evidence, i.e., if we could have no evidence for such a thing, why would we ever claim it to be there in the first place?¹⁸

Nevertheless, such a response may still be easier to believe than Parfit's own solution that our identity may occasionally be indeterminate. As Bernard Williams writes, "To be told that a future situation is a borderline one for its being myself that is hurt, that it is conceptually undecidable whether it will be me or not, is something which, it seems, I can do nothing with; because, in particular, it seems to have no comprehensible representation in my expectations and the emotions that go with them."¹⁹ And later: "There seems to be an obstinate bafflement to mirroring in my expectations a situation in which it is conceptually undecidable whether I occur."²⁰ While Williams is more concerned with the normative issues surrounding my anticipation/concern about my possible future survival (as opposed to an exclusive focus on whether or not I will survive in the sorts of cases under discussion), because the rational appropriateness of one or the other such attitude is directly parasitic on facts about who the survivor will be, the "bafflement" he admits to having about appropriate expectations seems straightforwardly transferable to bafflement about identity's being indeterminate. Consequently, if option (c) is just as hard to believe – just as baffling – as (b), Parfit's comments will be

insufficient to move us to become reductionists on the matter (on the assumption that we adhere to the "default" view that our identity is always determinate).

Nevertheless, there is a much more decisive way to eliminate the view of the Sharp Borderline Theorist (from here on out, "the SBT") in (b), namely, by showing that the position is actually *impossible* to believe (without contradiction or complete absurdity). To see why, we need to take a fairly detailed look at just what would be involved in the claim that there is some sharp borderline on the Combined Spectrum, before which I am the survivor and after which I am no longer the survivor. By providing a few possibilities, we will see that the SBT is in a world of trouble.

I begin, however, with a few caveats. First, in what follows I presuppose a widely-held, specific materialist tenet: physical changes (to the brain) will result in mental changes. A Cartesian would beg to differ, of course, but we have seen practical reasons to ignore this view. Second, I assume the SBT to hold that the person at the far end of the spectrum, a person constituted by fully 100% of the cloned body and brain cells of Garbo, is just GarboR. I take this to be a matter of common sense as well (it is also maintained by the reductionist). With these assumptions in hand, then, let us consider the following possible sharp borderlines in order: (1) I cease to exist at the 50% mark; (2) I cease to exist at some point *above* the 50% mark; and (3) I cease to exist at some point *below* the 50% mark.

First, let us suppose the SBT maintains that if the scientist were to flip 50 switches (and leave the remaining 50 switches unflipped), I would cease to exist and *also* that GarboR would begin to exist at that point as well. In other words (to make the scenario as clear as possible), on this supposition I would be the survivor if only 49% of my original cells/characteristics were destroyed and replaced, but I would not be the survivor – and GarboR *would* be the survivor – in the case in which the scientist flipped 50 switches.²¹

It seems this response is problematic, however, for the SBT would then have trouble maintaining that GarboR begins to exist at this point. Why? GarboR would *also* have only 50% of her body and brain cells intact now, and according to the borderline just given, it seems she

could not be in existence either. In other words, if X does not exist when 50% of X's original body and brain cells do not exist, then neither I nor GarboR could exist at this point.

This response is too quick, however. After all, the SBT might say, what matters may simply be who you are to start with. You remain that person until the crucial transition-point, and then you become a different person. The borderline of 50%, then, would be the borderline for *ceasing to be* the same person, and GarboR isn't necessarily governed by *that* borderline. In other words, the only principle supposed to this point (call it "Principle C-50") is that X *ceases to exist* when 50% of X's body and brain cells cease to exist, but this principle is irrelevant for GarboR; consequently, it is not yet obvious why she could not begin to exist at the 50% mark.

Unfortunately for the SBT, this will not do. To see why, consider the following two-stage scenario. Suppose the scientist has me in the chair and then flips 50 switches (the "first stage"). At this point, according to the SBT, the survivor is GarboR. But now suppose the scientist keeps this new person in the chair and flips the other 50 switches (the "second stage"); more specifically, the scientist now flips the 50 switches corresponding to the *remaining* 50% of GarboR's cells (those not drawn upon in the first stage). So the person who is now sitting in the chair has fully 100% of the cloned cells from Greta Garbo. Nevertheless, the SBT, given the previous response, is committed to saying now that GarboR is *not* the survivor, given that 50% of GarboR's (post-first stage) body and brain cells ceased to exist. In other words, the *ex hypothesi* GarboR (the 100%-of-Garbo's-original-cloned-cells person) is not GarboR (the 50%-of-Garbo's-original-cloned-cells person), a contradiction. If the scientist in the first stage had merely flipped all 100 of the switches, the SBT would readily affirm that the survivor was GarboR. There seems no reason available to the SBT to justify denying this conclusion as well in the two-stage version (especially if we suppose the second stage to take place immediately after the first).

Consequently, for the SBT coherently to maintain the 50% borderline, he/she must alter his/her position and claim that, at the 50% mark (when neither I nor GarboR exist), *someone else* pops into existence (call this third person "Shoebo"). But this move raises an intractable

difficulty as well. To this point all we have is the proposed Principle C-50, that X ceases to exist when 50% of X's body and brain cells are destroyed and replaced. But consider another two-stage scenario. Suppose once again that I am seated in the dreaded chair, but this time the scientist flips only 25 switches (perhaps because that is the most he can switch at once with his fingers, forearms, and elbows). So only 25% of my body and brain cells are destroyed and replaced with those of GarboR (stage one). Now the scientist immediately moves to the *next* set of 25 switches and flips them (stage two). According to Principle C-50, *I* am the survivor after stage two, for I was the survivor after stage one (having lost only 25% of my body/brain cells), and during stage two only 25% of *my* (the person sitting in the chair post-stage one) body and brain cells have been destroyed and replaced. Nevertheless, throughout these two stages a combined total of 50% of my *original* body and brain cells have been destroyed and replaced, so Principle C-50 also seems to imply that *Shoebo* would be the survivor. But if the principle implies both that I am the survivor and that I am *not* the survivor (i.e., Shoebo's the survivor), it must be false.²²

Nevertheless, this contradiction may be avoided by pointing to an ambiguity in Principle C-50. On the one hand, it might be taken to mean that X ceases to exist just when X loses 50% (or more) of X's original body/brain cells *period*, no matter the process or length of time it occurs. On the other hand, it might be taken to mean that X ceases to exist just when X loses 50% (or more) of X's original body/brain cells *all at once*. On the former interpretation, then, the survivor of the two-stage 25% process would be Shoebo. On the latter interpretation, the survivor of each stage would be me.

A sufficient reason for thinking the principle must involve the former interpretation is that the latter interpretation would allow for the (by assumption) false conclusion that I could actually be the survivor at the *far* end of the spectrum, even though that person would consist in all of Garbo's cloned cells. For suppose the scientist has a four-stage process of reaching the end of the spectrum: he sits me in the chair and then flips 25 switches, keeps that person (me, on this interpretation) in that chair and flips 25 of the 75 remaining switches, and then repeats the

process two more times. Insofar as there would be no stage in which more than 25% of my body/brain cells would be destroyed and replaced, I would have to be the survivor throughout. So again, I would have to be the survivor, even after 100% of my *original* body/brain cells had been destroyed and replaced (via the four-stage process) by the cloned Garbo cells. But insofar as the basic assumption of the SBT's position is that I am clearly *not* the survivor when 100% of my body and brain cells have been destroyed and replaced, this interpretation of Principle C-50 cannot viably be maintained. Consequently, we must go with the first interpretation, according to which X ceases to exist just when X loses 50% or more of X's original body/brain cells *period*.

The problem now, however, is figuring out just when GarboR could possibly pop into existence. Principle C-50, while specifying when someone ceases to exist, tells us nothing yet about when someone *begins* to exist, i.e., how much of one's body and brain cells must be "activated" for one coherently to be said to exist. Given the SBT's proposed borderline on the spectrum, GarboR cannot coherently be said to pop into existence at the 50% mark. But then what percentage of her body/brain cells must be activated for her to be the person in the chair? As already mentioned, the SBT agrees that if 100 switches are flipped at once (where the process begins with me in the chair), the survivor is GarboR. But if Shoebo pops into existence at the 50% mark, the SBT has a new borderline problem with which to contend, viz., what is the precise borderline among the range of 50 possible switches on the far end of the spectrum before which Shoebo exists and after which GarboR exists? In other words, if Shoebo would be the survivor upon the flipping of 50 switches, and GarboR would be the survivor upon the flipping of all 100 switches, what precise point along the spectrum (between 50 and 100 switches) constitutes the boundary between Shoebo and Garbo? If 51 switches were originally flipped, would the survivor still be Shoebo? What if 99 switches were flipped?

Both possibilities once again reveal trouble for the SBT. Suppose the scientist once again runs a two-stage process. I begin in the chair and the scientist flips 51 switches. The survivor, we are supposing, would be Shoebo. But then (in the second stage) the scientist flips the remaining 49 switches. According to Principle C-50, Shoebo's the survivor. He/she lost only

49% of his/her original body/brain cells during this stage, so he/she has not yet ceased to exist. But once more, we've got a person with 100% of the original cloned Garbo cells sitting in the chair now, and by hypothesis any person with 100% of those cloned Garbo cells indeed *is* GarboR, so given this possibility, Shoebo cannot be the survivor at the 51% mark. Nor could he/she be the survivor at the 99% mark, for the obvious reason that the scientist could, in the second stage, flip the remaining switch, and we would have the same problem on our hands.

Might, then, someone *else* pop into existence at the 51% mark, someone other than me, Shoebo, or GarboR? No. For if we are still maintaining the specified interpretation of Principle C-50, we could run the two-stage process on this new person and once more get the same contradiction: the person at the far end of the spectrum is both GarboR (*ex hypothesi*) and not GarboR (because of the two-stage process in which the survivor after 51 switches are flipped would also be the survivor if the remaining 49 switches were then flipped).

What all of this means, then, is that for Principle C-50 to be the SBT's boundary-drawing principle for the Combined Spectrum, it must be filled in as follows: (a) X ceases to exist only if 50% of X's original body/brain cells cease to exist, regardless of the period/method of their destruction; (b) if exactly 50 switches are flipped, Shoebo pops into existence; and (c) if 51 or more switches are flipped, GarboR pops into existence. But this final attempt at specifying C-50 fails as well, given yet another version of the two-stage process. Suppose in the first stage, the scientist flips 50 switches. According to (a) and (b), I have ceased to exist and Shoebo pops into existence. Then suppose the scientist flips any of the remaining switches between 51 and 99 (the second stage). According to (a), Shoebo will be the survivor; according to (c), GarboR will be the survivor. Once more we have arrived at a contradiction, and Principle C-50 must be abandoned as untenable.

We must then try a different mark on the spectrum as a possible boundary for the SBT. Suppose the SBT were to claim that 51 switches have to be flipped (in the original Combined Spectrum case) before I cease to exist. In other words, the SBT would here say that I actually do

survive if 50% of my body and brain cells are destroyed and replaced, but I fail to survive if 51% are destroyed and replaced. Call this Principle C-51.

Once more there is an ambiguity in the principle that needs to be cleaned up. For the same reason as with Principle C-50, we should interpret the principle as meaning that X ceases to exist just when 51% (or more) of X's original body/brain cells are destroyed and replaced, *regardless of the period/process involved in their destruction*. This interpretation prevents the SBT theorist who holds this principle from falling prey to another two-stage variation. If the principle is interpreted as meaning that X ceases to exist upon having 51% of X's body/brain cells destroyed and replaced *period*, then if the scientist were to flip only 50 switches (stage one), and then were to flip the remaining 50 switches (stage two), I would be the survivor once more at the far end, which has repeatedly been ruled out as a possibility.

On this view, then, if the scientist flips any number of switches over 50, no matter the number of stages in which such flipping takes place, I cease to exist. Who, then, is it that *does* exist at that point? It must be that the survivor either is or is not GarboR. Let us suppose first that it is not GarboR. Since the survivor cannot be me either, we'll once more call the survivor Shoebo. But this move will not work, for the standard two-stage reasons. If the scientist flips 51 switches (the first stage), then (we're supposing) Shoebo pops into existence. If the scientist then flips the remaining 49 switches, on Principle C-51, the survivor would still be Shoebo. But by hypothesis the person existing at that point would have to be GarboR. To avoid this contradiction, then, we must suppose that the person who pops into existence at any point on the spectrum from 51 on would have to be GarboR. Will *this* move work?

Notice first that if the SBT goes with this interpretation of Principle C-51, he/she is no longer susceptible to problems stemming from two-stage variations of the case. Once the 51% mark is hit, regardless of how many stages it takes to do so, I cease to exist and GarboR begins to exist. So a move that plagued all variations of Principle C-50 is avoided here. There remains, however, a serious problem with Principle C-51. To see why, instead of considering a *two-stage*

version of the Combined Spectrum, let us consider two different *versions* of the Combined Spectrum as given.

First, suppose I am seated in the chair and the scientist flips 50 switches. According to the SBT adhering to (the suitably specified) Principle C-51, I am the survivor. So far, so good. But now consider a different version of the Combined Spectrum, one in which GarboR begins in the chair, i.e., the scientist takes his replica of Greta Garbo from the bank of her cloned cells and starts her off in the Spectrum chair. Now suppose the scientist has all of *my* cells available to him to replace any of those he destroys from GarboR. He then flips 50 switches, destroying 50% of GarboR's body and brain cells and replacing them with 50% of my corresponding body and brain cells. Now the SBT would have to say that, according to Principle C-51, GarboR survives such a transformation (after all, 50% of her original body and brain cells are still intact).

But now note what the SBT is committed to: flipping fifty switches makes the survivor either me or not me, *depending on who was sitting in the chair at the beginning of the process*. In other words, if we start with *me* in the chair, and we destroy and replace 50% of my body and brain cells with 50% of GarboR's, I'm the survivor. But if we start with GarboR in the chair and replace 50% of her cells with mine, *she's* the survivor. Yet the survivor in both cases would have exactly the same percentage of cells originating with GarboR and me, i.e., in *both* cases, the survivor would have 50% of my original cells and 50% of GarboR's original cells. What, then, could possibly distinguish the two survivors? Again, it could only be that origins matter, i.e., what distinguishes the two survivors is their *history*. In the first version the survivor is me because it was me originally in the chair and the alteration was identity-preserving. In the second version the survivor is GarboR because it was her originally in the chair and the alteration was identity-preserving. Nevertheless, this rejoinder yields yet another contradiction.

To see why, consider one last bizarre twist on an already bizarre case. Suppose first that I am seated in the special chair and the entire *left* side of my body is zapped out of existence, while the remaining right side is carefully preserved in the scientist's cryo-chamber. Then the restored Garbo Replica is seated in the same chair and the entire *right* side of her body is

destroyed, with the remaining left side preserved in the cryo-chamber. According to Principle C-51, both GarboR and I will be the survivors once our missing parts are replaced. But now comes the fun part: the scientist then removes both halves from the cryo-chamber and fuses them together. Seated now in the chair is a person with 50% of my body and brain cells and 50% of GarboR's body and brain cells. And both of the fused halves started this long strange trip in the same chair.

Now the SBT adhering to C-51 is forced into a contradiction, for he must maintain that the fused survivor here is simultaneously both me *and* GarboR, i.e., the survivor is both me and not me, it is both GarboR and not GarboR. Who began the switch-flipping process in the chair is now seen as what it is: an irrelevant consideration, one that helps the SBT not one whit here. Consequently, this position yields yet another contradiction.

A last-gasp reply for the SBT theorist might run as follows: "The fusion here described is a fusion of you and GarboR, which (*ex hypothesi*) is supposed to be thoroughly symmetrical with respect to your status versus hers and candidates for identity with the resulting person. Given this symmetry, the resulting person is not identical to *either* of you, since you're not identical to each other, and you both were equally good candidates for being identical with the resulting person. Consequently, the fusion brings into existence a new person, and there is no contradiction yielded."²³

Unfortunately, this response is utterly ad hoc, and it quite simply denies Principle C-51, which directly implies that one's identity is *preserved* when only 50% of one's original body/brain cells are destroyed and replaced. If the sharp borderline is at 51% (or higher), then, quite simply, I survive at the 50% mark. And insofar as the same must hold for GarboR, the fused person in this scenario must be both me and not me, both GarboR and not GarboR, which *is* a contradiction. It is also important to note that this same argument applies if the SBT picks *any* point on the Spectrum between the 50% and 100% marks, for at each such instance we could show that by running this same fusion scenario we get a contradictory answer regarding the identity of the survivor at the 50% mark.

Let us then consider the final option for the SBT, viz., the borderline is somewhere *below* the 50% mark. Unfortunately, with this option we again have the possibility of Shoebo popping into existence and confounding the determination of identity all the way throughout the Spectrum, given the possibility of two- (or n-) stage variations. For example, suppose 25% is chosen as the boundary after which I cease to exist. To avoid the possibility of me being the survivor all the way to the end of the Spectrum (via a four-stage process in which the scientist flips only 25 switches at a time), we must again suppose that a loss of 26% (or more) of my original body/brain cells, *period*, ends my existence. But now we need a specification for the point at which GarboR pops into existence. If it is at the 26% mark, then a two-stage process in which the scientist first flips 26 switches (first stage) and then flips the remaining 74 switches (second stage) would yield a contradiction: the survivor at the end (the person with 100% of Garbo's cloned cells) would be both GarboR (by hypothesis) and not GarboR (because the survivor after the first stage was GarboR, who then would have to have ceased to exist after the second stage, given that during that stage more than 25% of her body/brain cells were destroyed and replaced). So once again we would have to introduce Shoebo as the survivor at the 26% stage. And once we have done so, no matter where we mark the boundary for where GarboR begins to exist, we will run into a contradiction. Suppose it is only after 100 switches have been flipped. Then if the scientist runs a four-stage process, flipping 26 switches the first three times and only 22 at the final stage, the survivor at the final stage would have to be both GarboR and not GarboR. For after the first stage (26 switches flipped), the survivor would be Shoebo. After the second stage (the next 26 switches flipped), the survivor would have to be someone new, call it Shoebo2. After the third stage (the next 26 switches flipped; a total of 78 switches flipped), the survivor would be Shoebo3. And upon the flipping of the final 22 switches, the survivor would have to be both Shoebo3 (whose identity was preserved throughout the fourth stage) and GarboR (by hypothesis). And again, such contradictions can be yielded if *any* point on the first half of the Spectrum is picked by the SBT as the borderline for my ceasing to exist and *any* point after that is chosen as the borderline for when GarboR begins to exist, *with one exception*.

That exception is the 1% mark. Suppose, then, that the SBT takes this final position – call it Principle C-1 – according to which I cease to exist just when 1% (or more) of my original body/brain cells cease to exist. On this view, GarboR could not begin to exist until the 100% mark, for if she popped into existence at any earlier stage, a two-stage process in which at least one more switch were flipped would kill her, a possibility which would force the SBT into a contradiction.²⁴ So the only viable way for the SBT to go here is to maintain that at every point on the spectrum, a different Shoebo pops into existence. If one switch is flipped, Shoebo1 pops into existence; if two switches are flipped, Shoebo2 pops into existence. And so forth, unless all 100 switches are flipped, in which case at that point GarboR pops into existence.

In short, the SBT is forced to admit that I cease to exist when 1% of my body/brain cells cease to exist. But even *this* is not entirely accurate, for there is certainly a range of alterations possible between the 0% and 1% marks on the Spectrum. Suppose, for instance, the scientist set up a "Mini-Combined Spectrum," where one switch flipped destroys and replaces only .01% of my body/brain cells, and 100 switches flipped destroys and replaces 1% of my body/brain cells. If the SBT holds the person at the 1% mark to be Shoebo1 (i.e., not me), then by the blitz of arguments given previously, the SBT would be forced to admit that the person at the .01% mark is not me either (it would be Shoebo.01, say). And between the 0% and .01% marks on the Mini-Spectrum, there is yet another range of alterations possible. And so on.

In other words, the SBT is ultimately forced by this reasoning to admit that I cease to exist just when *any change whatsoever* occurs to my body or brain cells. If just one subatomic particle, anywhere on my person, is lost or altered, I die. In short, I (the person who wrote the word "I") am now dead. But such a conclusion is utterly absurd.

Thus there simply *is* no sharp borderline on the Combined Spectrum to which one could point without contradiction or absurdity. Consequently, the second proposed reaction to this case – the claim that there is some sharp borderline somewhere on the Spectrum, before which the survivor is me and after which the survivor is not me – fails, and we are left with the final, reductionist response – that there simply is no determinate answer to the question of my survival

in the middle range of the Spectrum – as now the *only* coherent option. Personal identity may be indeterminate. And as hard as this may be to believe, it remains much easier than either of the other two contradictory or deeply absurd replies.

Remember how we got started here. The materialist non-reductionist must maintain the following conditional: if we are separately existing entities, then questions of identity will always have determinate answers. But now we can run modus tollens: it is not the case that questions of identity will always have determinate answers. Thus, we are not separately existing entities. Reductionism thus wins by default.

Even though reductionism is true, however, that does not mean we yet have answers to our motivating questions. To get such answers requires further specification of *which* of the variety of possible reductionist views is most plausible. Nevertheless, what I have tried to show herein is that such work may now proceed without being haunted by the specter of either non-reductionist souls – now seen as practically irrelevant – or other separately existing entities – now seen as either logically impossible or hopelessly absurd.²⁵

NOTES

1 Joseph Butler, "Of Personal Identity," in John Perry, ed., Personal Identity (Berkeley: University of California Press, 1975), p. 99. "Future state" may be ambiguous, however, perhaps meaning, for example, merely *tomorrow*, as well as perhaps meaning an afterlife of some sort. At any rate, whether or not Butler meant the afterlife does not diminish that fact that most people led to the issue of personal identity are driven by that meaning of "future state."

2 There might seem to be exceptions, of course, but these are not as clear-cut as some believe. For example, we may hold parents morally responsible for the actions of their child. Nevertheless, what we might really be doing in such a case is holding the parents responsible for some past action (or inaction) of *theirs*, which led directly or indirectly to the child's doing what he/she did.

³ There are actually two other methodologies available here. The first is to take our commitments in these practical arenas as given, as brute data without need of justification (or as already justified), and then figure out a theory of identity than can account for them all, i.e., a theory answering the question "How are such commitments *possible*?" Another alternative is to start with certain of our fairly settled, considered convictions about some aspects of our lives, and then proceed to reorganize the various implicit principles involved in these convictions into a coherent theory of identity. From there, we would explore how the theory might apply to other, less clear, aspects of our lives. In engaging in this process, though we treat neither our initial convictions nor our theory as beyond revision. They would merely be "provisional fixed points," and they would each be subject to adjustment given the twin demands of consistency and intuitive appeal. This latter is of course the method of "reflective equilibrium," and it is a method I actually favor in the arena of personal identity. Nevertheless, because the methodology sketched in the text is the one most often employed in this arena, and because it yields a conclusion compatible with what I believe we would conclude anyway with the reflective

equilibrium approach (although I certainly will not argue for this point herein), I will jump on the bandwagon and employ it myself in this paper.

4 Derek Parfit, Reasons and Persons (Oxford: Oxford University Press, 1984), p. 210.

5 See *ibid.*, p. 207.

6 Eric Olson suggests that most versions of the Physical Criterion are actually versions of the Psychological Criterion, insofar as they point to the continuity of the brain (or parts thereof) as crucial to identity only because of its role in the person's psychology. See his The Human Animal: Personal Identity Without Psychology (Oxford: Oxford University Press, 1997), pp. 16-21.

7 I should point out that the Cartesian Ego view is not the only possible non-reductionist View. The Cartesian Ego view holds that we are essentially separately existing entities, entities that are distinct from our brains, bodies and experiences, and it further holds that these separately existing entities are purely mental entities. A different version of this non-reductionist View, however, might hold that we are separately existing physical entities, "of a kind that is not yet recognised in the theories of contemporary physics" (Parfit, p. 210). A third kind of non-reductionism would agree with reductionism that we are not *separately existing* entities, but would still maintain that personal identity does not consist in facts about physical or psychological continuity; that there is still some further fact(s) involved. Parfit calls this version of non-reductionism the "Further Fact View" (p. 210). Nevertheless, all three versions of non-reductionism hold that personal identity consists in some further fact(s) other than those facts about brains, bodies and experiences. And all three versions claim that questions of identity always have determinate answers. (See Parfit, pp. 239-243 for a discussion of this last point.)

8 *Ibid.*, p. 227.

9 *Ibid.*, pp. 227-228.

10 Reprinted in Joel Feinberg and Russ Shafer-Landau, eds., Reason & Responsibility, 11th edition (Belmont, CA: Wadsworth Publishing Co., 2002), pp. 434-454.

11 Ibid., p. 439.

12 Parfit, pp. 236-243.

13 The question of whether or not these are two stages of the *same* person is precisely what is at issue.

14 E.g., p. 228.

15 Again, see Parfit, p. 228.

16 Ibid., p. 229.

17 Ibid. Incidentally, it is rather odd for Parfit to appeal to our commonsense reactions to this case, given that he ultimately argues that our commonsense reactions to such cases are generally *wrong*.

18 Ibid.

19 Bernard Williams, "The Self and the Future," in John Perry, ed., Personal Identity (Berkeley: University of California Press, 1975), p.193.

20 Ibid., p. 196.

21 To make things even clearer, I want to emphasize that I am preserving Parfit's original set-up of the scenario, according to which there are a range of cases in each of which there would be a single change to the person in the chair, as opposed to saying that the Spectrum involves a single case in which there would be a series of changes. So I am not envisioning a case wherein the scientist flips 49 switches, say, *and then flips one more*; instead, I am envisioning two different cases: one in which the scientist flips 49 switches, and another in which the scientist flips 50 switches. In my "Selves and Moral Units" (Pacific Philosophical Quarterly, v. 80, no. 4 (December 1999): 391-419), I offer a Revised Combined Spectrum in which changes to the subject *are* introduced gradually (to make a very different point), but here I am making use of Parfit's thought experiment as given.

22 This variation on the thought experiment obviously borrows from Reid's Brave Officer counterexample to Locke.

23 I am grateful to Terry Horgan for suggesting this reply.

24 For example, if in the first stage 1 switch was flipped, and the SBT claimed that GarboR was the survivor, then if the scientist (in the second stage) flipped all 99 remaining switches, she would both be the survivor (by hypothesis) and not the survivor (because she would have ceased to exist after losing at least 1% during the second stage after having been the survivor of the first stage).

²⁵ I am grateful to Terry Horgan and John Tienson for their very insightful comments on an early draft of this paper.