Personal Identity and Ethics
A *Brief Introduction*

David Shoemaker

BROADVIEW GUIDES to PHILOSOPHY

Library and Archives Canada Cataloguing in Publication

Shoemaker, David, 1964-
Personal identity and ethics : a brief introduction / David Shoemaker.

Includes bibliographical references and index.
ISBN 978-1-55111-882-6

1. Self (Philosophy). 2. Identity (Philosophical concept). 3. Ethics.
I. Title.

BD450.S45 2008 126 C2008-904149-6

Broadview Press is an independent, international publishing house, incorporated in 1985. Broadview believes in shared ownership, both with its employees and with the general public; since the year 2000 Broadview shares have traded publicly on the Toronto Venture Exchange under the symbol BDP.

We welcome comments and suggestions regarding any aspect of our publications—please feel free to contact us at the addresses below or at broadview@broadviewpress.com.

North America
PO Box 1243, Peterborough, Ontario, Canada K9J 7H5
2215 Kenmore Ave., Buffalo, NY, USA 14207
Tel: (705) 743-8990; Fax: (705) 743-8353
email: customerservice@broadviewpress.com

UK, Ireland, and continental Europe
NBN International, Estover Road, Plymouth, UK PL6 7PY
Tel: 44 (0) 1752 202300; Fax: 44 (0) 1752 202330
email: enquiries@nbninternational.com

Australia and New Zealand
UNIREPS, University of New South Wales
Sydney, NSW, Australia 2052
Tel: 61 2 9664 0999; Fax: 61 2 9664 5420
email: info.press@unsw.edu.au

www.broadviewpress.com

This book is printed on paper containing 100% post-consumer fibre.

PRINTED IN CANADA

CHAPTER ONE

Personal Identity and Immortality

Gretchen Weirob, a philosophy professor, has gotten into a terrible motorcycle accident, and she now finds herself in the hospital with only a day or two left to live. Despite being close to death, however, she is lucid, and is thus able to carry on an extended conversation with her two friends, Sam Miller, a chaplain, and Dave Cohen, her former student. Weirob is a lifelong atheist, but as her death approaches, she wonders about the possibility of immortality, and yearns, as many of us would, for the comforts of being able to anticipate surviving the death of her body. Through the next three evenings, right up until her death, the three friends discuss the nature of personal identity and immortality, with Miller and Cohen trying desperately to find a way to provide Weirob with the comfort she seeks (within the demanding strictures of reason), but to no avail: Weirob dies believing that there is simply no way for her to survive the death of her body, and thus no reason to anticipate immortality.

This is the "plot" of John Perry's imaginative *A Dialogue on Personal Identity and Immortality*,[1] and insofar as it is a terrific introduction both to the most historically influential theories of personal identity as well as to the motivation many people have for becoming interested in personal identity in the first place—worrying about the possibility of life after death—we will take it as our initial guide in this chapter. Along the way,

1 Bibliographical information on this and other writings referred to in this book will be found at the end of chapters.

we will critically evaluate the various theories discussed in the dialogue, as well as a few variations the dialogue participants overlook. By the end of the chapter, we will see not only how difficult it is to come up with a coherent criterion of personal identity that allows for the possibility of immortality, but also how difficult it is to come up with a coherent criterion of personal identity at all.

BACKGROUND

Weirob wants the comfort of being able rationally to anticipate surviving the death of her body. What is involved in this sort of rational anticipation, though? There are two elements. First, it cannot be rational to anticipate the occurrence of something that's just impossible. So in order for it to be rational for Weirob to anticipate surviving her body's death, it must at least be possible for her to survive her body's death. Further, this is all that Weirob demands: she is not asking whether or not she will definitely survive her body's death, nor is she asking whether or not such survival is probable. Instead, she simply wants to know if it's *possible* to survive, that is, if it's conceivable without contradiction or serious absurdity (one might think of this as **metaphysical possibility**). This is a very minimal constraint, it would seem, although as we will see it's actually a constraint that turns out to be very tough to meet in this case.

The second element Weirob assumes is that personal identity is a necessary condition of **rational anticipation**. What does she mean by this, though? In general, to anticipate something is to look forward to it. So I may anticipate the end of the current war in Iraq, say, or I may, as in the old commercial, anticipate the ketchup's finally coming out of the bottle onto my hot dog. But Weirob has in mind a very specific form of anticipation that involves *looking forward to actually having certain experiences as occurring "from the inside."* Think here of what it's like to remember some recent experience: you have a representation of a past you lived through, so you relive the sights, sounds, and even smells of what it was like to actually undergo that experience from the inside, as

the experiencer of that event. Anticipation is just this aspect of remembrance cast into the future: to anticipate some experience is thus to have an imagined representation about what some experience one expects to have will feel like from the inside.

Furthermore, to *rationally* anticipate some future experience, for Weirob, is to do so in a way that is correct, or that makes sense. So suppose Cohen and Miller could establish not just the *possibility* of heaven, a "place" where there are lots of happy people communing with God for all eternity, but heaven's actual existence. This wouldn't yet be enough for Weirob to rationally anticipate anything: she wants it to be possible, not only that there will be persons existing in an afterlife setting, but that one of those persons *will be her*. After all, how could it be rational for her to anticipate the experiences of a stranger in heaven? Indeed, even if that stranger were exactly like her in every way, if that person weren't in fact *her*, then how could she rationally look forward to the experiences that that person would undergo in heaven? Instead, it seems what's necessary is that, for such anticipation to be rational, it must be possible for there to exist someone in heaven who is identical to—who is the same person as—Weirob on Earth.

In general, someone has the burden of proof in an argument if the claim that that person is advancing is not obviously true. Weirob holds that the claim "survival of death is possible" is certainly not obviously true, and she makes explicit that her dialogue partners have the burden of proof here by reiterating an uncontroversial fact: her body will eventually cease to exist. And we can make this even more explicitly true by stipulating that her body will be cremated immediately after she dies. Given this fact, she asks, how could *I* still exist? Now this way of formulating the question indicates that she holds a **materialist** conception of the "I": it is *physical*, consisting of matter. In putting the challenge in this way, Weirob also gives us our first criterion of personal identity, what she takes to be the default view that Miller and Cohen have the burden of replacing:

> **The Body Criterion**: *X at t1 is the same person as Y at t2 if and only if X's body is the same as Y's body.*

While we will discuss the subtleties of this view later, for now it should be obvious that, if it's true, and if the rationality of anticipation depends on personal identity, and if one's body does indeed cease to exist after death, then it would be irrational to anticipate surviving the death of your body, because such survival would be *impossible*. After all, if your body is destroyed upon your death, then no one could have your same body after that death, and so no later person could possibly be you—even if both God and heaven exist.[1]

The gauntlet Weirob throws down to Miller and Cohen, then, is to show her that things could possibly be otherwise. More specifically, she presents

> **Weirob's Challenge**: *come up with an alternative criterion of personal identity that (a) could provide a means, a mechanism, to enable Weirob to rationally anticipate surviving the death of her body, and (b) does not yield a contradiction or deep absurdity, that is, it's actually possible.*

A theory fulfilling both of these conditions would thus allow her rationally to anticipate the afterlife, giving her the comfort for which she is so desperate in her final days, and, not inconsequentially, giving each of *us* some reason to hope that there's more to our own lives than this merely mortal coil.

---

1 Peter van Inwagen has actually concocted a scenario, however, in which immortality is possible (and thus it could be rational to anticipate survival in the afterlife), even assuming the truth of the Body Criterion. The way in which he accomplishes this is actually by denying Weirob's so-called uncontroversial fact. He maintains it is possible that, just as you are about to die, God whisks your body to heaven and replaces it on earth with an exact replica that then dies in your place. This would make the person's body in heaven thus continuous with–the same thing as–your body on earth, and so would make that heavenly person you. See van Inwagen, *The Possibility of Resurrection and Other Essays in Christian Apologetics* (Boulder, CO: Westview Press, 1998), Chapter Three, "The Possibility of Resurrection," pp. 45-51. While this scenario might indeed be metaphysically possible, it remains unclear just what this process of "whisking" involves, and so it remains unclear just what would make that body in heaven the "same" as the body on earth. It would also, more disturbingly, turn God into a deceiver, someone who allows us to think that we and our loved ones will die, when in fact all those who actually die are imposters.

The Soul Criterion

After some initial misunderstandings about the nature of the challenge, the chaplain Sam Miller offers a familiar and expected alternative criterion:

> **The Soul Criterion**: *X at t1 is the same person as* Y *at t2 if and only if X's soul is the same as* Y*'s soul.*

How would this answer Weirob's challenge? The soul is ostensibly your essence, and something that is different from, and can exist independently of, your body. Thus, the soul could provide a means to enable Weirob to rationally anticipate surviving the death of her body in the following way: her soul could continue to exist after her body dies, and perhaps be transplanted into another body in the afterlife. And so, more generally, if the soul is your essence, and a soul doesn't have to die along with your body, then *you* could continue living after that body dies. The Soul Criterion thus seems to meet the first demand of Weirob's Challenge. What about the second demand, though? Is it actually possible for things to work this way?

Before answering this question, we have to do some basic philosophical spadework, that is, we have to get clear on just what we're talking about here. What exactly *is* a soul, after all? It turns out this is a rather vexed question, one that has yielded many different sorts of answers throughout history. Consider just the two most influential answers. Plato took the soul to be what a person really was, which he thought was an essentially nonphysical thing, so that persons were the incorporeal occupants, perhaps even the prisoners, of their bodies. Aristotle, by contrast, took the soul to be merely the formal design, the organizing principle, of a living body. Persons on this conception, therefore, are like coins, whose essence consists in both their formal design (shared by all coins) and their particular physical manifestations.

Now obviously whether or not you can establish the possibility of immortality by means of souls will depend on what you think the nature of

souls actually is. If you accept Aristotle's view, then the soul just isn't a substance, a *thing*, that could ever even exist independently of a particular living body, so once your body dies, its soul would have to be no more (destroyed coins simply have no design). This is why, for example, Thomas Aquinas, a Catholic theologian and philosopher who accepted Aristotle's conception of the soul, insisted on the resurrection of the *body* in the afterlife, for without their bodies, persons could not survive their deaths. So it looks like the only conception of the soul that has a chance of enabling immortality is Plato's (a conception also shared, more or less, by Descartes).

But even on this conception of the soul as purely incorporeal, some questions remain. What is its nature, after all? Is it a purely psychological substance—a thing whose whole essence is to think (as Descartes maintained)—or is it a substance that, while *having* a psychology, is in principle separable from it? Furthermore, is a soul something I *have* (distinct from me), something I *am*, or something else entirely? These are very hard questions, and it's quite unclear how to answer them. Indeed, once we allow that the soul would have to be incorporeal, we are more or less resigned to the problem of having no direct and reliable way of determining whether or not they exist or what they'd be like, given that we can directly and reliably know about the existence and nature of only those substances we can experience with our senses. But then, if we cannot directly determine the existence or nature of souls, how could they play any real role in personal identity, which often involves precisely such direct and reliable judgments?

Indeed, Weirob raises a version of this problem herself in the following argument (**Weirob's Soul *Reductio***[1]):

1. If the Soul Criterion were true, we could never have the grounds to judge that X is the same person as Y.
2. Sometimes we *do* have the grounds to judge that X is the same person as Y.

3. Thus, the Soul Criterion is false.

1 The term "*reductio*" here is short for *reductio ad absurdum*, a Latin phrase meaning, roughly, "reduces to absurdity." Running a *reductio*, then, involves showing that someone's argument has implications that are either contradictory or absurd, which itself implies that their original argument is either false or implausible.

In other words, if we believe, as Miller does, that sameness of incorporeal souls is what preserves persons' identities across time, but we have no direct way of reidentifying souls (given that we can't sense them), then we would also have no direct way of reidentifying persons either. But surely this is false, given that we directly reidentify people all the time and seem to do so for very good reasons. When I return home every evening and greet the person standing there with a kiss, I'm judging that this person is my wife, the same person I married and the same person I kissed goodbye in the morning. But if what made the morning and the evening person identical were just their identical souls, I would have absolutely no direct and reliable way to determine that they in fact were the same person (and so perhaps I should be withholding those kisses!). This can't be right, though. These sorts of common, immediate, everyday reidentifications must have rational grounds, if anything does.[1] Thus, any theory that implies that these sorts of ordinary reidentifications are groundless *must* itself be false, given that it implies a contradiction with the facts of ordinary identity attributions. The Soul Criterion, it seems, cannot pass the possibility condition of Weirob's Challenge.

How might a defender of the Soul Criterion respond? What one needs to do is find some intermediate link between the grounds we use in ordinary judgments of reidentification and the identity of those incorporeal souls, and there might be two ways to do so: (a) via bodies, and (b) via psychology. The first response might go as follows. Ordinarily, we reidentify people visually, by reference to their bodies. But what bodies might do is simply serve as an *indicator* of the soul "inside," such that same body implies same soul. We thus infer the existence of the same soul upon seeing the same body. Ultimately, then, what preserves your identity across time is your ongoing soul, and while we cannot reidentify souls directly, we can reidentify them *indirectly* via their bodies, which always carry within them the same soul.

---

1 This is not to say that I couldn't be *mistaken* in my reidentifications. If my wife has an identical twin, they might occasionally play a trick on me by switching places, in which case I might judge that some person is my wife when in fact she isn't. This would be a case in which I was mistaken about what my grounds for reidentification actually were, however—they have to consist in more than mere physical similarity, it would seem—not a case in which there were no such grounds.

Unfortunately, this won't do, for how could we ever establish a principle like "same body, same soul" in the first place? In other words, why think that there's a one-to-one correlation, or *any* direct correlation, between bodies and souls at all? In driving home this point, Weirob offers Miller a box of chocolates, and he picks out one that has a swirl on top, given that it indicates the presence of caramel inside. Miller, then, operates on the principle "same swirl, same filling." The question, though, is how he could ever have established a principle like that, and the answer is that he's observed both sides of the equation: he's actually been able to bite into a swirl-on-top chocolate repeatedly, and he has then *tasted* the caramel inside each time. But this method simply won't work to establish the principle "same body, same soul," given that we can never directly experience both sides of the equation, that is, we can never *taste* a soul. So it's just impossible for us to *establish* the correlation Miller wants between bodies and souls.

The second attempt to link ordinary reidentification to the identity of souls comes via the intermediary of *states of mind*. Ordinarily, of course, we reidentify people via their bodies. But bodies alone are merely *indicators* of identity, and although usually reliable, they could mislead. If the person with the body of your best friend suddenly started talking and behaving precisely in the manner of your worst enemy, or like a complete stranger (even failing to recognize you), you might begin to have doubts that she was indeed the same person as you knew before, despite your judgment that this person has the same body as your friend. So while our grounds for reidentifying people make reference to their bodies, this is just a shorthand way of reidentifying their psychologies, their states of mind. And given that states of mind are simply states of *soul*, according to Miller, reidentifying someone's psychology is a reliable and indirect method of reidentifying her soul. Consequently, what makes X and Y at different times the same person is the presence of the same soul in both, and what enables us to reidentify X as Y is the sameness of psychological characteristics, which themselves bear a one-to-one correlation to souls, and which are themselves (typically) reidentified via reidentification of bodies.

This last attempt fails as well, however, as Weirob demonstrates with a

discussion of rivers. Suppose an expert on a few local rivers could reidentify them solely on the basis of the state of their water. So some rivers have cloudy, brownish water, while some rivers are crystal clear. One might think, then, that the expert reidentifies the rivers according to the principle "same water, same river." But of course, the water the expert points to in reidentifying some river isn't the *same* water he's seen before—that's all long downstream. Instead, it's just *similar* water, or in similar states to the water he's seen before. So the same river at different times consists in different, albeit similar, water.

Analogously, then, the same person could consist in different, albeit similar, souls (or minds). It's perfectly possible, after all, that similar (but distinct) states of mind are attached to similar (but distinct) souls, which would be, remember, substances we couldn't see, touch, smell, taste, or hear. But because it's impossible for us to reidentify souls directly, any number of hypotheses about their relation to me is fair game: I could indeed have had one soul attached to this body and psychology since birth (or before), but I might also have gotten a new, exactly similar soul during my mid-life crisis, or every year on my birthday, or even have had a constant river of exactly similar souls flowing through me. Notice, then, how correct judgments of personal identity would have to depend on which of these scenarios occurred, if the Soul Criterion were true, but because we cannot establish any clear linkage between bodies, psychologies, and souls, we cannot ever know if our judgments of personal identity are correct. But again, this seems clearly false, given our confidence in the grounds of the many reidentifications we make in our ordinary lives. The overall conclusion, then, is that the Soul Criterion, while meeting the first part of Weirob's Challenge, cannot meet the second part: it does not provide a possible mechanism to get her (or us) to the afterlife, given the contradiction it yields with respect to our ordinary judgments of identity.

What are we to make of this overall argument, Weirob's Soul *Reductio*? Notice first that Weirob does not deny the existence of souls. Instead, she grants that they might indeed exist, but insists that, due to their allegedly incorporeal nature, they simply couldn't play any role in our judgments of identity. This way of putting it, though, reveals that the Soul

Criterion may not be false after all. Indeed, what Weirob seems to be doing is confusing the two senses of "criterion" discussed in our introductory chapter, the metaphysical sense and the epistemological sense. The Soul Criterion, as given, is a purely metaphysical criterion, purporting to explain what *makes* X and Y identical. Weirob's objection, though, is epistemological, complaining about how we could never *know* that X and Y are identical if the Soul Criterion were true. But for a defender of the Soul Criterion, such an objection might well be irrelevant, for souls could still constitute the identity-preservers for persons, even if we had no means of tracking their trajectories through space-time. This would constitute a straightforward rejection of the second premise of Weirob's Soul *Reductio* (which maintains that we do indeed have the grounds to make judgments of reidentification), and it would allow that it still might be possible for me to survive the death of my body via my soul.

Still, there remains something compelling about Weirob's objection. She claims that she hasn't based her "argument on there being no immaterial souls..., but merely on their total irrelevance to questions of personal identity, and so to questions of personal survival,"[1] but this isn't quite right, as we've just seen. Souls may be *quite* relevant to the question of personal identity itself. What they're not relevant to, however, are the *practical concerns* we have that are related to personal identity. In other words, all of the prudential and moral cases discussed in the Introduction that seem to depend on personal identity actually presuppose that we *can* make correct judgments about when that identity relation obtains. Holding people responsible, compensating them, determining the moral relation between fetuses and the adult humans into which they develop, determining the moral relation between early- and late-stage Alzheimer's patients, and (what's most relevant to the present chapter) rationally anticipating some future experience(s)—all of these practical concerns and commitments presuppose our ability to identify and track whatever criterion of identity turns out to ground them; they presuppose a tight connection, that is,

1 Perry, "A Dialogue on Personal Identity and Immortality," in Joel Feinberg and Russ Shafer-Landau, eds., *Reason and Responsibility*, 12th edition (Belmont, CA: Wadsworth/Thomson Learning, 2005), p. 371.

between the metaphysical and epistemological senses of "criterion of personal identity." Consequently, any theory of personal identity to which we lack this kind of epistemological access is just going to be *practically* irrelevant.

And this is precisely the case with the Soul Criterion. *It could be true*, of course. But given that we in fact reidentify people via their bodies or their psychologies (what else could we do, either reliably or directly?), and given that souls would have no necessary connection to *either* (as Weirob correctly points out), it's very difficult to see what the point of appealing to souls in a metaphysical criterion of identity could possibly be. So we can either (a) allow the truth of the Soul Criterion, in which case we have to allow both that there's a disconnection between the nature of personal identity and our practical concerns, and also that our reidentification practices are likely ungrounded and potentially wildly mistaken, or (b) we can insist on a tight connection between the nature of personal identity and our practical concerns, and thus reject any theory of personal identity—like the Soul Criterion—that denies this connection. Because (a) would have wildly unsettling implications for many aspects of our daily lives, there are good practical reasons for adopting (b), and thus rejecting the Soul Criterion.

Note what we both have and have not done here. There are many compelling arguments that have been given to deny either the existence or the coherence of souls. Obviously, if these arguments succeed, then the Soul Criterion is false—if identity is ever preserved across time, it couldn't be because of some non-existent substance. But we have not appealed to these sorts of arguments. Instead, we have suggested that, even if there are souls, they aren't relevant to the *practical* questions we are asking, and so any criterion appealing to them just misses the point of the general inquiry. Is this a satisfactory dismissal of the Soul Criterion? This is the sort of very abstract matter we will take up in the final chapter. But for now, it's important to see how the Soul Criterion fails to meet Weirob's Challenge in a different way than Weirob herself thinks it fails: while Weirob thought the soul could have been a mechanism warranting rational anticipation of survival but the criterion of identity appealing to the soul couldn't possibly

be true, we have suggested instead that while such a criterion could in fact be true, the soul couldn't actually be a mechanism warranting rational anticipation of survival. Imagine, for instance, that there could be someone in heaven with my soul, but who had neither a body nor a psychology anything like mine, someone who didn't remember my life or experiences at all. What possible reason could I have to anticipate his experiences, *even if he is me*? It would be no different from my anticipating the experiences of a complete stranger, and so its possibility would surely fail to provide Weirob (and us) with the comforts of anticipation being sought.

Nevertheless, it may still be rational to anticipate the possibility of surviving death, even without reference to souls. How so? One might appeal to a criterion of identity that is more closely connected to our ordinary practices of reidentification but that is in principle separable from one's body. And that's just what Miller tries to do in the second night of the dialogue.

The Memory Criteria

In arguing that the Soul Criterion fails, Weirob makes reference to the ordinary practice of reidentifying other people: we simply make no reference to souls when engaged in that everyday practice. Nevertheless, third-person reidentification isn't the only sort of reidentification we engage in; another kind is *first-person*. So when you groggily and gradually wake up in the morning, you know who you are—you reidentify yourself, for example, by thinking that *you* should have gone to bed earlier last night, or getting angry with yourself for having forgotten to study for today's exam when *you* had a chance yesterday. Now what are your grounds for such first-person reidentification? Do you check to see if your soul is the same as that of the person who got into bed so late last night? Of course not, but we already knew the Soul Criterion was a dead end. But here's the kicker: you also don't check to see if your *body* is the same as that of the person who got into your bed last night.

In fact, you can imagine waking up in a totally different body, yet still remaining "yourself." Indeed, this sort of thought is familiar from a vari-

ety of literary and cinematic flights of fancy. The opening line of Franz Kafka's *The Metamorphosis*, for example, is as follows: "One morning, as Gregor Samsa was waking up from anxious dreams, he discovered that in bed he had been changed into a monstrous verminous bug."[1] Notice that our reaction is not that this is incoherent, or utterly incomprehensible. Instead, while acknowledging the exceedingly unusual nature of the metamorphosis, we can still grant that it's *conceivable* and, for all we know, perfectly possible. This is also true of the numerous popular "body swapping" movies made over the years, such as *Big*, *Vice Versa*, and *Freaky Friday*. All of them assume that it's possible for a person's identity to be preserved (and known about first-personally) regardless of any particular body that person might have. If all of this is true (and Weirob is skeptical), then there simply is no *substance* underlying personal identity, that is, it necessarily consists in neither souls nor bodies. Instead, it consists in various relations among the various stages of persons, relations to which we have access from the first-person standpoint.

To illustrate, suppose you attend a baseball doubleheader (two baseball games played back to back), and you get up in the middle of the seventh inning of the first game, during a one-to-one tie, to buy a hot dog. As it turns out, the line is fairly long, and you don't get back to your seat for another hour. Upon your return, you see that the score is one-to-one, but you ask the person next to you, "Is this the same game I was watching when I left?" It's equally possible, after all, that the first game ran long or that the second game already started and the teams are once again tied. Now what would identity of games consist in? That is, what is the nature of the question you have asked? Are you asking about *souls*, in some sense, about whether or not this game has the same "spiritual essence" as the game you'd been watching? Clearly not: we certainly don't think that anything of the sort underlies a baseball game. Are you instead asking about *bodies*, in some sense, about whether or not these are the same players or the same field as before? No. It's possible, for instance, that the same game could proceed without *any* of the same players and on a different field.

1 Franz Kafka, *The Metamorphosis*, trans. by Ian Johnston. Available as e-text on the web, at http://www.mala.bc.ca/~Johnstoi/stories/kafka-E.htm.

This would be the case, say, if there were an earthquake in the fourth inning of a World Series game that stopped play and damaged the stadium, such that they had to complete the game in the opposing teams' stadium, where both managers, in an attempt to fire up their teams, replaced all the starting players with the benchwarmers. What, then, are you asking about with your question?

You're asking about the *internal relations* of the game, about how the parts of the game—its various events, like strikes, hits, outs, runs, and so forth—relate to the *whole* game. What you're asking is whether the out one team has just recorded counts as an out in the first game or the second. A baseball game, as a whole, is simply comprised of all its various individual events, and as long as the parts are connected *in the right way*, then it's still the same game. What counts as the right way? Well, that's to ask for an actual criterion of identity for baseball games, and to answer, we would have to make detailed reference to the rules: there are nine regularly scheduled innings (although sometimes more are played if there's a tie, and sometimes only eight and a half are played when the home team's ahead), and each of these innings is comprised of three outs per team, and some outs consist in catches of fly balls, and other outs consist in a batter swinging at and missing a pitch three times, and so forth. But at any rate, when one baseball-event is related to various other baseball-events in the right way, given the rules of the game, they are all parts of the *same* game.

What Miller thus suggests to Weirob is that a person is like a baseball game: a person, as a whole, is simply comprised of all its related parts, and is not some underlying substance, such as a soul or a body. What thus allows you to know who you are in the morning is not your reidentification of some persisting substance, but is instead your awareness of the relation that connects your various parts into a single whole, the true relation of personal identity. But what exactly is this relation? It is clearly psychological, and Miller borrows from the seventeenth-century philosopher John Locke to make it more specific: what unites our various stages are *memories*. We can start afresh, then, with a new theory of identity based on this insight:

Memory Criterion #1 (MC1): *X at t1 is the same person as Y at t2 if and only if Y remembers the thoughts and experiences of X.*

Could this criterion meet Weirob's Challenge? On its face, it easily meets the first demand. It is certainly possible that 1000 years from now there will be some person in heaven who remembers Weirob's life (from the "inside"), thinking of Weirob's past as her own, in the way we all do upon waking up in the morning. If she did this, she would, on this criterion, *be* Weirob. And just as it's rational for me now to anticipate the thoughts and experiences of the person who will wake up in my bed tomorrow and who will remember my thoughts and experiences today, so too it would be rational for Weirob to anticipate the thoughts and experiences of this person in heaven who will remember her life on earth. This would, in effect, be like Weirob's finishing her "baseball game" on a different field.

Can MC1 meet the second demand, though? Could the criterion be true without implying any contradictions? As it stands, the criterion runs into two immediate difficulties, noticed by two critics of Locke's original view in the century after its publication. First, as Joseph Butler pointed out in 1736, MC1 implies that if I cannot remember some past experience, then that experiencer could not be me. In other words, I have existed, on this criterion, only during those moments I now remember. But this must be false: just because I don't remember having lunch last Thursday doesn't mean that none of last Thursday's lunch-havers were me! Surely identity can persist through some memory loss.[1]

A related, but potentially more devastating, problem came from

1 One thing to consider is that this point is an objection to the *necessary condition* of MC1, claiming that my remembering some past experience isn't necessary for making that past experiencer me: he may be me even if I *don't* remember his experiences. All Weirob really wants, though, is a criterion that provides a *sufficient condition* for identity, identifying some substance or relation that, if present, will ensure her survival. So she would be perfectly happy if a memory of some past experience was enough to guarantee that the rememberer was identical to the experiencer, even if memory wasn't necessary for identity, for that could still allow her a mechanism to survive the death of her body: as long as someone in heaven remembered Weirob's life and experiences, that person would be Weirob. Nevertheless, our interests are wider than Weirob's, for we'll ultimately want a full-fledged criterion of identity that explains what it *always* requires, and for this task the objection above is relevant.

Thomas Reid in 1785, from the **Brave Officer Case**. Suppose that, at 10, a boy steals some apples from a neighbor's orchard, and then at 40, as a brave officer, he steals the enemy's flag in battle, and then finally, at 80, he's a retired general. Furthermore, suppose that, as the 40-year-old is stealing the flag, he fondly remembers stealing the apples as a 10-year-old, and as the 80-year-old is relaxing in his rocking chair, ruminating about his life, he fondly remembers stealing the enemy's flag, but—and here's the troublemaking part—he has no memories whatsoever of stealing the apples. What does Locke's view say about the relation between the retired general and the apple-stealing kid? Because the brave officer (BO) remembers the experiences of the apple-stealer (AS), the BO is the same person as the AS. And because the retired general (RG) remembers the experiences of the BO, the RG is the same person as the BO. But then if AS = BO, and BO = RG, then AS = RG; in other words, logic demands that, given Locke's theory, RG is the same person as AS. Nevertheless, RG doesn't remember any of the experiences of AS, so Locke's theory also implies that RG is *not* the same person as AS. Consequently, Locke's theory implies a contradiction—RG both is and is not identical to AS—and so the theory itself cannot be true.

Both problems, however, can be resolved with a fairly easy patch-up job, for all we need to do is amend MC1 by having it appeal to an overlapping chain of direct memories, rather than to a direct memory link itself. In other words:

> **Memory Criterion #1a (MC1a):** *X at t1 is the same person as Y at t2 if and only if Y directly remembers the thoughts and experiences of* X, *or* Y *directly remembers the thoughts and experiences of some Z, who directly remembers the thoughts and experiences of some Q (who remembers R, who remembers S, who remembers T, as needed)...who directly remembers the thoughts and experiences of* X.

I remember what my yesterday's self did, and he remembers what *his* yesterday's self did, and so on, back to last Thursday's lunch-haver. So while I now have no direct memories of what I had for lunch last Thursday, I'm

connected via this chain of memories to someone who *does* remember, and this is all that's needed to respond to the first objection. Furthermore, this amendment can also easily deal with the Brave Officer Case, for now the RG will be identical to the AS, given the chain of overlapping memories between them, even though the RG has no direct memories of the AS's experiences, so the contradiction is dissolved.

Nevertheless, there is another serious problem: how does MC1a distinguish between *seeming* to remember and *actually* remembering? Consider, for example, the psychiatric patient who thinks he's Napoleon: he seems to remember fighting the Battle of Waterloo and sleeping with Josephine. He is obviously deluded. But now notice the problem for MC1a as it pertains to immortality: suppose there's a person in heaven 1000 years from now who seems to remember my thoughts and experiences. What's to prevent this from being just like the Napoleon case, one in which there's someone who is *deluded* into thinking he's me?

Any successful memory criterion of personal identity will obviously have to refer only to *genuine* memories. What's needed, then, is a way to distinguish between genuine memory and merely apparent memory: what makes one person's memories genuine, after all, and another's merely apparent? Suppose, then, that Y at t2 seems to remember the experiences of X at t1. One might very well be inclined to say that Y's memories are genuine if and only if Y *actually had* the experiences he now remembers. Indeed, how could I have genuine memories of anything other than *my own* experiences; isn't this just what the nature of memory consists in?

Unfortunately, this is a deeply problematic response, for it renders the overall enterprise circular, and so undermines the establishment of MC1a. Here's why. Suppose we've got a person in heaven at t2 who claims to be Weirob. What would make this person Weirob? On MC1a, she is Weirob if and only if she remembers Weirob's thoughts and experiences. Now suppose this heavenly person does indeed seem to remember Weirob's life. It remains possible, given the Napoleon case, that she could be deluded, so we now have to ask, "What makes her memories genuine?" Well, goes the possible response above, she was the person who actually had those experiences she now remembers, that is, *she was*

Weirob. But now the problem should be obvious: what makes her Weirob? She remembers Weirob's life. What makes her memories of Weirob's life genuine? She was Weirob. But what makes her Weirob? And round and round we go.... The problem is that it looks as if genuine memories *presuppose* identity, that you cannot have genuine memories of someone else, that memories by their very nature (when genuine) *reveal* your own past to you. But if that's the case, memories cannot constitute the identity relation, for in order for my memories of some past experiences to be genuine, I *already* have to be identical with that past experiencer.

Nonetheless, is it necessarily the case that I can have genuine memories of only my own experiences? It may not be. Consider the following possibility (drawn from arguments given by Sydney Shoemaker and Derek Parfit). Suppose scientists develop a way to copy a memory trace of your European vacation into me, such that I seem to remember standing underneath the Eiffel Tower (even though I have never been to France). Further, suppose I know that I've never been to France, but I know you have, and I further know that the scientists have performed this procedure on me. My memory would thus not be a delusion (for I wouldn't think *I* was the one who'd been to Paris), nor would it be of an experience that happened to me. Why couldn't it thus count as a genuine memory, an accurate memory of some experience that nevertheless didn't presuppose identity? If so, then what makes it a genuine memory would be a purely causal matter: the memory must be caused by *the experience* that I now remember. Genuine memories thus simply have to have an orthodox causal history. They must be caused by an experience of the remembered event and so be a product of the ordinary causal chain: an experience causes a trace in the brain, a trace which is then later tapped into in one's remembrance of the experience. What makes a memory merely apparent, then, is that it wasn't caused in the right way, that is, it wasn't caused by the experience that's being remembered. So, for example, the Napoleon guy may seem to remember fighting the battle of Waterloo, and while that memory will have a cause (perhaps a trauma in childhood), its cause will not be the *experience* of fighting in the battle of Waterloo, rendering it merely apparent.

These remarks thus yield a new version of the Memory Criterion:

> **Memory Criterion #2 (MC2):** *X at t1 is the same person as Y at t2 if and only if (a) Y seems to remember the thoughts and experiences of X (either directly or via an overlapping chain of memories), and (b) Y's seeming to remember is caused in the right way.*

Can this criterion thus pass Weirob's Challenge?

In articulating the view, we have suggested that for Y's memory to be caused in the right way, it must be one that has been stored in Y's brain. But if that's the case, then MC2 can't pass the first part of Weirob's Challenge, for it wouldn't seem to provide a mechanism for immortality. After all, if my brain is what houses my genuine memories, and it is destroyed along with the rest of my body upon my death, then I simply couldn't survive the death of that body. So while there could be a person who exists 1000 years from now in heaven who seems to remember my life, he couldn't be me, given that his "memories" wouldn't have an orthodox causal history. They would have been caused by God, not by memory traces preserved in the same brain.

Nevertheless, why think that sameness of brain matters here? After all, what we really want in demanding genuine memories is that *the storage of information be reliable*, not that the process of storage take the same route every time. So normally, of course, an experience occurs, which causes a trace in the brain, and the experience is later retrieved from that brain as memory. But it seems the process could just as well go as follows: an experience occurs, which causes a trace in your earth-brain, the information of which is downloaded by God upon your death onto a Divine Flash Drive (DFD), which is then plugged into a new body in heaven, the information is uploaded into the new brain, and this newly activated person in heaven now remembers the experiences of the earth-you. As long as God and the DFD are reliable, why should we care that the process of transfer is a bit out of the ordinary? Sure, there may be no God, or if there is a God he may not care about helping us survive our deaths. But it's *possible* that God exists, and that God cares, and that God could store the information from our memories on earth reliably. And given these possibilities, we would have a mechanism to get us from here to there.

Is this really possible, however? Can the criterion relying on this mechanism avoid contradictions? As it turns out, it can't, and in seeing why we will run across a very famous and puzzling thought experiment, one that we will return to many times throughout the book. It is a case of *fission*, and one way to illustrate it comes from consideration of a possible heavenly case, what we will call **Divine Duplication**. Suppose that, upon Weirob's death, God did the downloading process described above, and then uploaded all of her memories into a newly created body's brain in heaven. This person thus "wakes up" thinking she is Weirob, that she has survived her death, and according to MC2, she'd be right: her memories would be caused in the right way, namely, via reliable information storage. Call this person Weirob 1.0. But now suppose God likes the results so much that he creates another new body and plugs the DFD into it, uploading Weirob's memories again. Now MC2 directly implies that this person—call her Weirob 2.0—is identical to the original Weirob as well. But if both versions 1.0 and 2.0 are identical to Weirob, then, by the transitivity[1] of identity, they must be identical to *each other*. But clearly they are not: they each occupy different locations in space-time, for one thing, and they could go on to lead very different lives after this, maybe even going on to reside in opposite sides of heaven. To insist they are the *same* person is simply to stretch the concept of personhood into something unrecognizable.

What we have, then, is the following problem for MC2:

1. If MC2 is true, then the products of the Divine Duplication of Weirob would be identical with one another (by transitivity).
2. Both resulting people would *not* be identical with one another (given their different locations in space-time, for one).
3. Thus, MC2 is false.

---

1 *Transitivity* is a property of certain relations, such that if A has that relation to B, and B has it to C, then it must be that A has it to C. An example of a transitive relation is *bigger than*: if A is bigger than B, and B is bigger than C, then it must be that A is bigger than C. An example of a relation that is not transitive is *is a friend of*. It could be that A is a friend of B, and B is a friend of C, but A is not a friend of C. It's often assumed that identity must be transitive: after all, if A is identical with B, and B is identical with C, then A must be identical with C.

The logical implication of MC2 cannot be true, so neither can MC2. It thus fails to pass the second part of Weirob's Challenge.

There is one remaining reply, however. What got MC2 into trouble was the duplication. If God made only *one* person in heaven with Weirob's memories, then that person would be Weirob, and it would be possible for her to anticipate surviving the death of her body. And it is perfectly possible that God makes only one version of each of us in heaven. So what we can do is simply stipulate that, where one copy is made, the original survives, and where any other number of copies is made, the original doesn't survive:

> **Memory Criterion #3 (MC3):** *X at t1 is the same person as Y at t2 if and only if (a) Y seems to remember the thoughts and experiences of X (either directly or via an overlapping chain of memories), (b) Y's seeming to remember is caused in the right way (via any reliable cause), and (c) no other beings satisfy conditions (a) and (b).*

Think of this as a **No Competitors** version of the Memory Criterion: as long as there's no competition, no other person whose memories of your life are genuine, you have survived. If there's no one in existence with genuine memories of your life, or if there exists more than one person with genuine memories of your life, you have not survived.

Unfortunately, as Weirob makes clear, this last-ditch attempt fails to meet the second part of her challenge as well, for it is deeply absurd. To see why, suppose that God is a bit of bumbler, and that sometimes his newly created bodies survive and sometimes they don't. Suppose, then, that he creates a new body in heaven upon Weirob's death on earth, and he uploads her memories into this new body. According to MC3, this person is Weirob. But then suppose God makes another body and uploads Weirob's memories into this person. Now neither person is Weirob, *even though the first person was Weirob for a day*. But this is just too weird.

Consider yourself, after all, from the first product's point of view. "I made it!" you might think, "I've been brought back to life by a kind and loving God. It's great to be alive and to be *me*, good old Weirob." But then

suppose God creates the duplicate. Now you, the person who *was* Weirob, are no longer Weirob. Suddenly, you'd have to think "Well, I *was* Weirob, but now I'm someone else, a deluded imposter, someone who thinks she's Weirob but isn't." But now suppose the duplicate simply collapses and dies (it wasn't one of God's better efforts). Now what might you think? "Yes, I'm back! It's great to be me—Weirob—again!" But then suppose God resuscitates the duplicate. "*!$%#*, now I'm no longer Weirob, but just a deluded nobody once more!" And so forth.

The problem is that whether or not you're the same person would change as new competitors pop into existence. But this is terribly absurd. Surely my identity doesn't depend on what happens to other people! After all, it's perfectly possible that God has at this moment made a duplicate of me in heaven, or on the opposite side of the universe, or in Albuquerque. But then MC3 would imply that I have suddenly ceased to exist—even though I've undergone no physical or psychological changes at all—and at the moment of duplication I would have been replaced by an entirely new person, one who *thinks* he's Shoemaker, and has "memories" of Shoemaker's life, but who is just terribly deluded. But this cannot be right. Whether or not some future person is me must instead depend solely on the relations between us, not on the existence or non-existence of other people. And so, while we have found no straightforward contradiction here, we have found a deep absurdity, which, as Weirob says, has the same weight as a contradiction, and so this third twist on the memory criterion fails to pass the second part of her challenge.

The Body Criterion

Am I then just my body? If so, then immortality is impossible. And this is certainly what Weirob continues to believe, right up until her death, despite the best efforts of her friends to convince her otherwise. Once we start to examine this view closely, however, it too becomes difficult to believe.

Begin with the case of *Who is Julia?*, a work of fiction that's treated as if it were real in Perry's dialogue. The set up is as follows: as Julia

attempts to save a young child, she is run over, and her body is left mangled, although her brain is fine. Meanwhile, the child's mother, Mary Frances, has a brain aneurysm at the scene, although her body is fine. Both people are rushed to the same hospital, where the remarkable Dr. Matthews is able to transplant Julia's healthy brain into Mary Frances's healthy body, resulting in one healthy person with Mary Frances's body and memories of Julia's life. What, then, has happened to Julia (and Mary Frances, for that matter)?

The general method behind the presentation of this sort of puzzle case (and we'll run across several others) is to identify and draw upon your intuitions[1] in various interesting cases, and then use those intuitions as the data for which a successful theory of personal identity must account. In this respect, then, a theory of identity will resemble a scientific theory, and, as in the science case, the better theories will be assessed as such, in part, by virtue of their explanatory power, that is, insofar as they are able to explain more of the data in a more adequate fashion. The *Julia* case is important to this end, as we will see, because it prizes apart two features that are ordinarily wedded together—our bodies and our brains—and it asks, "If they were separated, where would *I* go?" So what are your intuitions in this case? Who is the survivor: Julia, Mary Frances, or someone else?

Most people believe that Julia is the survivor: the resulting person would at least have apparent memories of Julia's life, would believe she was Julia, would recognize the people in Julia's life and fail to recognize those in Mary Frances's life, and so forth. Furthermore, there's no reason to believe that her memories are delusions, given that they would have been caused in the right way—the ordinary way—by having been preserved in and recalled from the same brain in which the experience was originally recorded.

Weirob, of course, disagrees, maintaining that the survivor is Mary Frances, a *deluded* Mary Frances, true enough, but Mary Frances nonetheless. Perhaps some plausibility for this position can come from consideration of other cases of organ transplants. If my liver were to fail and the

1 *Intuitions* are common-sense judgments that people would make prior to considering philosophical arguments and theories.

surgeons were to replace it with a donor liver, say, we would all agree that I was the survivor of that operation, that I'd just gotten a liver transplant. And if my heart were to fail and the surgeons were to replace it with a donor heart, we would certainly say once more that *I* was the survivor, that I'd gotten a heart transplant. Why not, then, say the same as well in the Mary Frances case, that *she* was the survivor, and that she'd just gotten a brain transplant?

But the brain isn't like those other organs, many of us want to say. It alone, after all, is what preserves our memories, and our psychology generally, and so for that reason it's in a very different category from the heart and the liver, at least with respect to *personal* identity. Sure, you might think, organs like the heart and the lungs keep me alive, but it's the psychology provided by the brain that is the *me* being kept alive, and while I could be kept alive by *any old* set of heart, lungs, and liver (and so could survive in Mary Frances's body), the me that's being kept alive is a psychological being, and that's what my uniquely important brain preserves.

If this is the intuition most of us share in the *Julia* case, and if such intuitions count as appropriate sorts of data, then we would have a powerful reason to reject the Body Criterion, for it would thus yield the wrong answer in this case. We would also have a condition any plausible alternative criterion would have to meet, namely, that it account for our intuitions that personal identity depends in some way on brain-based psychological relations.

Nevertheless, this isn't a knock-down objection to the Body Criterion, in part because some may share Weirob's intuitions on the case, and in part because there are, as we will see, seriously unclear or implausible features of the brain-based psychology view. Are there, then, other, less controversial, objections one might raise? There are two.

First, consider the real life case of the Hensel twins, Abigail and Brittany. (*See photo.*[1]) They are conjoined twins, sharing their internal organs below the waist, while having two spinal cords, two hearts, three lungs, and two stomachs. Nevertheless, they have only two arms and two legs,

1 Photo from http://www.search.com/reference/Abigail_and_Brittany_Hensel.

and their nervous systems are connected and partially shared, which allows them to coordinate their activities fairly well. They can run, ride a bike, swim, and play the piano.

What are we to say here? Given our ordinary conception of "body," it seems clear that Abigail and Brittany share a single body: they have one torso, after all, and they have two arms and two legs. How many persons are there, however? Clearly, there are two. The twins repeatedly stress that they are two distinct individuals, and it would be nearly impossible to deny that this is the case. Each writes separately in her own hand, they disagree with one another, they take pride in their individual accomplishments, and so forth. But if, as Weirob would have it, X and Y are one and the same person if they have the same body, and Abigail and Brittany have the same body, then Abigail and Brittany *are one and the same person*, which is clearly false. In this actual case, the Body Criterion gives the wrong answer, and so it is deeply flawed. (As we will see in the next chapter, the Biological Criterion, which is similar in certain respects to the Body Criterion, might have a way to deal with this case.)

A second problem comes from a more far-fetched case offered by Derek Parfit. An advocate of the Body Criterion must admit that your identity could be preserved through the loss and replacement of one of your fingernails. Similarly, it could be preserved via the replacement of your finger, your hand, your kidneys, and your heart. But there must be *some* point after which your identity would no longer be preserved, for if everything but your brain were replaced (as happens to Julia), Weirob insists you would no longer exist (or you would have gone wherever your body now was). But where precisely is the line at which your identity would be lost, before which the survivor would be you, and after which the survivor would be someone else? Is it at the 50% mark, so that if 49% of your body were replaced, you would survive, but if 50% were replaced, you wouldn't? Perhaps it's at the 51% mark?

We simply don't know. And, what's worse, we simply *could not* know. Even if we could actually do such large-scale replacement, and we could ask the person at the 50% mark who she is, why should we believe what she says? After all, the person with Julia's brain will say she's Julia, but an advocate of the Body Criterion wouldn't believe her; instead, this advocate would think that the survivor was a *deluded* Mary Frances. But then there would be no way in principle we could determine where the line was marking the difference between identity and non-identity, between life and death.

Nevertheless, there would have to be such a line. But then the deeper worry would be this: how could the difference between life and death consist simply in the replacement of a few cells? Suppose the 51% mark is the line in question (even though we could never know that), so that you'd be the survivor if only 50% of your body were replaced, but you wouldn't be the survivor if 51% of your body were replaced. This would make the difference between identity and non-identity, between life and death, depend on a very small amount of physical change: that 1% of bodily replacement would take with it your *entire* identity. But this is very hard to believe, especially in light of the fact that your identity would clearly be *preserved* through a 1% replacement at the *early* stage of the spectrum (getting a finger transplant, say).

The Body Criterion is thus in serious trouble. Sameness of body doesn't seem to be what preserves my identity across time. But there's a *part* of my body that may indeed do the trick, namely, my brain. Let us turn then, finally, to consideration of that possibility.

The Brain-Based Memory Criterion

Recall the insight that I can reidentify myself without reidentifying either my body or my soul (if I have one). This was the insight motivating the move from a substance-based criterion to a relational criterion, that is, the Memory Criterion. But as we saw, this view fell prey to the duplication objection: if there were two people in the future bearing the appropriate memory relation to me, according to this criterion they would both have to be me, and so by the transitivity of identity they would both have to be identical to one another, which they clearly would not be.

What got that view into trouble was its criterion of genuine memories, of what counts as a memory's being "caused in the right way." Because the advocates of the view in Perry's dialogue wanted to establish the possibility of life after death, they allowed that memories were like software and that memories would be genuine just in case the information they contained had been stored reliably.[1] This move allowed for God to download my earth memories onto the DFD, create a new body and brain in heaven, and then upload my memories into that new brain. But the metaphor of memories as software is what enabled the possibility of Divine Duplication, of that software being "uploaded" into two (or more) different "hard drives."

Nevertheless, this isn't the only way the view might go. Instead, we might think of memories in terms of a *hardware* metaphor, such that one's memories are genuine—caused in the right way—just in case they are

1 Of course, there may be good independent reasons to head in the direction Perry's dialogue participants do on this matter. One might, after all, simply be convinced by thinking about various cases that genuine memory is just a kind of accurate information storage and that any reliable vehicle for such storage would thus be sufficient to preserve identity. However, this position would be just as vulnerable to the duplication thought experiment as the one motivated by a desire to establish the possibility of immortality.

preserved *in the very same brain* in which the remembered experience was originally stored. What matters, on this view, is thus preservation of the same brain:

> **The Brain-Based Memory Criterion (BBMC)**: *X at t1 is the same person as Y at t2 if and only if Y seems to remember X's thoughts and experiences (either directly or via an overlapping chain of memories), and Y's memories are caused in the right way, namely, via the same brain.*

This view eliminates the possibility of immortality, of course, but for now we're simply trying to come up with a plausible criterion of personal identity to account for rational anticipation in everyday cases. We'll return to discuss immortality in a bit.

The advantages of BBMC should be obvious. For one thing, it allows us to account for the intuition most of us share that Julia would be the survivor in the *Who is Julia?* case. The survivor, remember, would seem to remember Julia's life and experiences, and she would have Julia's brain. A second advantage is that, although we lose immortality on this view, we at least avoid the troublesome Divine Duplication case (and its earthly variants), insofar as one and only one brain could preserve genuine memories. A third advantage is that this view easily accounts for our intuitions in the Hensel sisters case: Abigail and Brittany are two different persons insofar as there are two different streams of memory provided by two different brains.

This view nevertheless has some significant problems, unfortunately, and they're brought out by consideration of some science-fiction cases. Suppose, first, that teleportation were possible. That is, suppose it were possible for you to step into a machine on earth that scanned and recorded the state of all your cells while zapping your body and brain out of existence, and then faxed that information to Mars, where another machine created, from new matter, a body and brain exactly like yours. In other words, your body and brain would be destroyed on earth, and moments later someone exactly like you in every single respect would walk out of the machine on Mars. Suppose the machine were 100% reliable. Would you consider using it?

One way to think of the question is this: would this be a way of traveling, where *you* would be going from earth to Mars in a matter of seconds, or would it be a case of death, then duplication? Many people's intuitions lean towards the former (perhaps from years of watching teleportation on *Star Trek* and other science fiction shows). But obviously, if you think you survive teleportation, that it's just a really fast form of travel, then you don't accept BBMC.

A second, and related, case casting doubt on the criterion is one to which Weirob refers, called Brain Rejuvenation, which is just an earthly version of the Divine Duplication case. Suppose your brain had weak blood vessels but that it was possible to replace it with a rejuvenated version, that is, possible to make a new brain, out of human tissue, that would be your brain's exact duplicate—it would consist in all the same psychologically relevant states—and then to replace your aneurysm-liable brain with the healthy duplicate. The resulting person would be exactly like you psychologically, seeming to remember your life and experiences, and thinking that he or she is you. Would you have the operation? *Should* you have the operation?

Your answer here will likely depend on where you stand on the question of whether or not the survivor would be you. But if you're an advocate of BBMC, both possibilities cause serious trouble. First, suppose you believe the survivor would be you. You've now abandoned BBMC, allowing instead that identity *doesn't* depend on sameness of brain, that the memories of the post-rejuvenation person would be genuine, despite not having been stored in and retrieved from the same brain of the person who experienced them. And beyond abandoning BBMC, you've now opened yourself up to the very same worries about duplication we ran into before, in the software version of the Memory Criterion.

On the other hand, if you believe the survivor *wouldn't* be you, you're sticking to your guns, but now you have some new questions to answer. First, it is unclear why preservation of the *very same brain* is so crucial to the preservation of identity in this case. Why think that it is? The resulting person would be exactly like you in every way, and he or she would still have a good 95% of your original body. Why think that that specific three

pounds of gray matter is what makes the difference between your life and your death? Indeed, isn't it just the *memory* preserved by your brain that matters to your identity? Why, then, put so much weight on the mere *means* of preservation, as opposed to the *content* of the preservation? At the very least, more needs to be said to give us a positive reason for why the very brain that you have is so absolutely crucial to your identity.

Second, and more importantly, BBMC actually undermines the motivation for adopting a memory-based view in the first place. Recall Miller's point about everyday self-identification: we can wake up knowing who we are without having to reidentify any substances, either souls or bodies: I don't have to check to see if my body or soul is the same as those of the person who got into my bed the night before to know that that person was me. Miller took this fact about us to motivate a relational view of personal identity, arguing (alongside Locke) that what explains this fact about self-reidentification is the relation of memory one now stands in to the person (oneself) being reidentified.

But now suppose you agree to undergo the Brain Rejuvenation, just to see what it is like (your own brain is perfectly healthy, say), so you're now in the operating room, where the surgeons have already removed your brain (call it B_1) and they have it sitting next to the duplicate brain (B_2). Suddenly, one of the surgeons trips and falls into the table holding both brains, and they fall off the table with a squishy thud and then slide across the floor, bouncing into the wall and into each other. The surgeons rush over, pick them up, and sterilize them, but now they've lost track of which brain is which. After some quiet discussion, they agree simply to stick one of the brains into your cranium and stick the other into a different live body they happen to have in the next room, and they then agree never to speak of this again.

So now there will be a person who wakes up with your body and either B_1 or B_2 (call this person BrainyOne), and another person who wakes up with a different body and either B_1 or B_2 (call this person BrainyTwo). Now the original advantage of the Memory Criterion is lost: neither BrainyOne nor BrainyTwo will really know who he or she actually is. They'll be unable to engage in self-reidentification, the key capacity motivating a move

to the Memory Criterion in the first place. Of course, they'll both *seem* to remember your life and experiences, but only one of them will have *genuine* memories of that life, insofar as only one will have your original brain, but neither will know – indeed, *no one* will know – which person is right and which person is wrong. So things simply become much more mysterious on this view, and BBMC loses much of its luster thereby.

We have now examined the same four possibilities for what preserves personal identity across time as Weirob and her dialogue partners: two types of substance criteria—a physical substance (the Body Criterion) and a non-physical substance (the Soul Criterion)—and two types of relational criteria—a software-based Memory Criterion and a hardware-based Memory Criterion. And all four theories were found seriously wanting. Indeed, it's very hard to believe that any of these theories could come close to making sense of rational anticipation. What we need to do, then, is either radically revise one or more of these theories or consider a very different type of identity altogether. In the next two chapters, we will pursue both strategies. First, however, we should consider where we stand on the topic that got us into this mess, namely, the possibility of making it into an afterlife.

The Possibility of Immortality

What, then, are we to say about immortality? Is it possible? Does Weirob, or do any of us, have any reason to look forward to it, to be comforted by it? Unfortunately, the prospects don't look very good. We have examined two general theories claiming to provide a mechanism for surviving the deaths of our bodies—the Soul Criterion and the software-based Memory Criterion—but neither of them, in any of their variations, could meet Weirob's Challenge: they were either irrelevant with respect to rational anticipation or led to contradictions or deep absurdity.

Does this mean, then, that immortality is impossible, and that we all ought to give up on looking forward to it? Not exactly. Indeed, we need to be very careful and precise about what's been shown here. It will be useful

to remind ourselves of the basic conditions of the enterprise. According to Weirob, it is rational for me to anticipate the experiences of some heavenly person (HP) only if:

1. HP is me, that is, personal identity is a necessary condition of rational anticipation;
2. There is a criterion of personal identity accounting for why HP is me that yields no contradictions or absurdities; and
3. The mechanism of survival provided by this criterion of identity is (at least in part) what gives me a reason to anticipate the experiences of HP.

The Soul Criterion violates Condition 3: even if souls could possibly exist, and even if they could be the substances preserving one's identity across time, they would just be irrelevant with respect to rational anticipation. Given their incorporeal nature, we would have no direct means to track them or reidentify them, and so they could not be the basis for our direct judgments of identity, either in others' cases *or our own*. But if I could have no direct grounds to determine that some future person would be me, what reason would I possibly have to anticipate his experiences? Souls thus couldn't explain the rationality of *any* future anticipation, and so couldn't provide any reason to anticipate the afterlife.

One could, of course, simply deny Condition 3 and embrace the Soul Criterion. This wouldn't be a very promising move, however, for it would entail that a host of practices in which direct reidentification is essential—practices to which we are deeply committed, such as responsibility-attribution, compensation, and on-going personal relationships in general—are completely unfounded. But it is very hard to believe that these practices could be unfounded, that we're just *guessing* when we reidentify one another. Denial of Condition 3 would thus be far too radical for serious consideration.

Memory Criterion #1 (in both its variations) is incomplete, failing to incorporate a criterion of genuine memories. MC #2 and MC #3, however, while incorporating an account of genuine memories, nevertheless violate Condition 2 above: they cannot handle the possibility of Divine Duplication without either a contradiction or a deep absurdity. Of course, one could

deny Condition 2, but this would be even less promising than denying Condition 3. Surely the mechanism constituting our identity must be possible!

Are there any criteria of personal identity that can meet these three conditions? It seems not, at this point. Given the uncontroversial fact with which we started, that our bodies will eventually cease to exist, any criteria meeting the afterlife conditions would have to rely on an essentially non-physical mechanism to get us from here to there. But either this mechanism will be provided by some kind of non-physical *substance* (with no necessary attachment to one's individual psychology)—a variation on the Soul Criterion—in which case it will violate Condition 3, or it will be provided by some kind of (non-physical) psychological software *relation*—a variation on the software-based Memory Criterion—in which case it will violate Condition 2 in dealing with Divine Duplication. In either case, then, it looks like what gets the view in trouble is its detachment from the earth-person's individual and unique *body*. But if viable theories of personal identity depend on the persistence of, or attachment to, that specific earth-body, then our prospects for immortality look particularly bleak.

There remains one last and extremely provocative response to the challenge, however, hinted at in Dave Cohen's final speech to Weirob. At this point in the dialogue, the participants have given up on the possibility of immortality, and they have focused instead on finding a way for Weirob to survive at least a few more years here on earth. As it turns out (in this fictional world), Dr. Matthews could perform a fascinating operation for her, transplanting her brain into another living body (a body currently without a brain). Weirob, however, refuses the operation, insisting that, insofar as it would be someone else's body, the resulting person would be someone else, despite the fact that that person would be, psychologically, exactly similar to Weirob. Cohen, unable to block Weirob's arguments for this view, makes one last appeal:

> Suppose you are right and we are wrong. But suppose these arguments had not occurred to you, and, sharing in our error, you had agreed to the operation. You anticipate the operation until it happens, thinking you will survive. You are happy. The survivor takes herself to be you, and thinks she made a decision before the operation

> which has now turned out to be right. She is happy. Your friends are happy. Who would be worse off, either before or after the operation?
>
> Suppose even that you realize identity would not be preserved by such an operation but have it done anyway, and as the time for the operation approaches, you go ahead and anticipate the experiences of the survivor. Where exactly is the mistake? Do you really have any less reason to care for the survivor than for yourself? Can mere identity of body, the lack of which alone keeps you from being her, mean that much? Perhaps we were wrong, after all, in focusing on identity as the necessary condition of anticipation....[1]

What Cohen is suggesting is that we should at least consider denying Condition 1 above. Perhaps what grounds rational anticipation isn't personal identity at all, but something else. And perhaps this something else is a mechanism that can survive the deaths of our bodies. If so, then while none of us can actually survive our deaths, it could at least be possible to rationally anticipate the experiences of some heavenly person, in the same way we may rationally anticipate the experiences of our future selves from day to day. But what could this "something else" possibly be? And do we really need to make such a radical move? Perhaps we can still get everything we want through a better-built theory of personal identity. In order to understand Cohen's suggestion and fully appreciate the motivation for it, then, we first need to analyze and evaluate the two most popular and sophisticated theories of personal identity around today, theories that have been designed to overcome many of the objections we have encountered to this point.

WORKS CITED OR REFERENCED IN THIS CHAPTER

Aquinas, Thomas. *Summa Theologica*. Translated by Fathers of the English Dominican Province. Benziger Bros. edition, 1947. Christian Classics Ethereal Library, http://www.ccel.org/ccel/aquinas/summa.html.

---

1 Perry, p. 383.

Aristotle. *On the Soul*. Translated by J.A. Smith. *The Internet Classics Archive*, http://classics.mit.edu/Aristotle/soul.html.

Butler, Joseph. "Of Personal Identity." In Perry, *Personal Identity*.

Descartes, Rene. *Meditations on First Philosophy*. http://oregonstate.edu/instruct/phl302/texts/descartes/meditations/meditations.html.

Harris, Barbara. *Who is Julia?* New York: Fawcett Books, 1977.

Kafka, Franz. *The Metamorphosis*. Translated by Ian Johnston. http://www.mala.bc.ca/~Johnstoi/stories/kafka-E.htm.

Locke, John. "Of Identity and Diversity." In Perry, *Personal Identity*.

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

Perry, John, ed. *Personal Identity*. Berkeley, CA: University of California Press, 1975.

——. "A Dialogue on Personal Identity and Immortality." In *Reason and Responsibility*, 12th edition, edited by Joel Feinberg and Russ Shafer-Landau. Belmont, CA: Wadsworth/Thomson Learning, 2005.

Plato. *Phaedo*. Translated by Benjamin Jowett. *The Internet Classics Archive*, http://classics.mit.edu/Plato/phaedo.html.

Reid, Thomas. "Of Mr. Locke's Account of Our Personal Identity." In Perry, *Personal Identity*.

Shoemaker, Sydney. "Persons and Their Pasts." *American Philosophical Quarterly* 7 (1970): 269-85.

van Inwagen, Peter. *The Possibility of Resurrection and Other Essays in Christian Apologetics*. Boulder, CO: Westview Press, 1998.

CHAPTER TWO

Personal Identity, Rational Anticipation, and Self-Concern

INTRODUCTION

Suppose that we persist in our commonsense belief that a necessary condition for my rational anticipation of some future person's experiences is that he will be me. After all, goes the normal thought, while I can imagine, be excited about, or have sympathy for the experiences of someone else, I can't *anticipate* that other person's experiences. So if anticipation is rational at all, it has to be (at least in part) because the person whose experiences I'm anticipating will be me.

The same, we might think, goes for the special type of concern known as *self-concern*. Suppose I find that my best friend will undergo torture this weekend. I will be very concerned about him. But if I find out that *I* will undergo torture this weekend, my concern is now of a different kind. What seems to justify this difference? It surely seems as if the justification has its source in the fact that in the second case the tortured person will be *me*, while in the first case he will not, which suggests that this sort of special concern is also grounded in personal identity.

If identity is really what grounds rational anticipation and self-concern, though, then we need to make a more valiant effort to see just what the right criterion of identity is in order fully to understand what implications it might have for these practical attitudes (as well as what it might imply about the possibility of immortality). In the previous chapter, we examined four such theories—Soul, Software-Based Memory, Brain (Hardware)-Based Memory, and Body Criteria—but saw that they were either false, seriously problematic, or just irrelevant to the issue of our practical concerns.

In this chapter, we will examine two much more sophisticated theories of personal identity, theories that, while similar in certain respects to the previous four, have nevertheless been designed to overcome many of the objections we launched against them. Most contemporary theorists support one or the other of the two theories we will examine here, and our job will be to see whether or not either one can truly avoid serious objections while also providing the grounding for anticipation and self-concern we have been seeking. We will also occasionally explore what implications they might have, if any, for the possibility of immortality. Let us begin, then, with the theory improving on the one many of you likely found most plausible in the first chapter.

The Psychological Criterion

Many of the problems we ran into regarding both the Body Criterion and the Brain-Based Memory Criterion (BBMC) were due to the difficulties both theories had in explaining our intuitive responses to certain cases, both real and science-fiction. So, for instance, the Body Criterion could not seem to deal with either the Hensel twins case or the *Who is Julia?* case. And while BBMC could handle both of those cases, it could not handle the teleportation and Brain Rejuvenation cases very well.

On the other hand, what seemed to handle all of these cases quite well was the software-based Memory Criterion, according to which personal identity is constituted by memory-relations, where memories are genuine insofar as they are caused by the remembered experience and the

memory-information has been preserved reliably across time. The fairly devastating problem for this theory, however, came from the possibility of Divine Duplication. So is there a way we can preserve the advantages of this theory, while avoiding the duplication fiasco?

There just might be. In fact, the most popular contemporary theory of personal identity, until quite recently, has been a robust expansion of the Memory Criterion that simply stipulates away the problem:

> **The Psychological Criterion**: *X at t1 is the same person as Y at t2 if and only if Y is uniquely psychologically continuous with* X.

There are several elements of the view to explain. First, this criterion appeals to psychological continuity, rather than mere continuity of memory. One reason should be fairly obvious: most of us would believe that someone's identity could be preserved through a bout of amnesia, that memory alone isn't the be-all and end-all of personal identity. What the Psychological Criterion does, then, is incorporate several other psychological relations that seem important to identity. There are four relevant relations that might obtain between X at t1 and Y at t2: (a) present-past relations, that is, Y now remembers the actions and experiences of X; (b) present-future relations, that is, X now intends to perform an action that Y later carries out in action; (c) persistence relations, that is, X has a belief, desire, or goal that persists across time to be held by Y; and (d) resemblance relations, that is, X and Y have very similar characters.

X and Y may bear these relations to one another either directly or indirectly. When they hold directly between the two person-stages—as when Y directly remembers an experience of X—call the relation one of *psychological connectedness*. Now across time there may be any degree of connectedness that obtains between two person-stages. So Y at t2 may remember only one experience of X's at t1 and that's it (i.e., Y has no other memories, intentions, beliefs, desires, or character elements of X). This would still mean psychological connectedness obtains between X and Y, albeit of the most minimal degree possible. Alternatively, Y may remember all of X's experiences, as well as carry out X's intentions, believe

everything X believed, desire everything X desired, and have a qualitatively similar character to X. Psychological connectedness of course obtains between them, but now to the highest degree possible. And one can imagine all sorts of degrees of connectedness in between that could obtain.

Of course, it's easy to imagine cases in which there are just *no* degrees of connectedness between two stages that are nevertheless stages of one and the same person. Consider the Brave Officer Case from Chapter One again, with a slight twist. Now instead of the eighty-year-old general not remembering any of the experiences of the 10-year-old apple stealer, suppose the retired general (RG) also bears no psychological connectedness of any kind to the apple stealer (AS). Nevertheless, we can still imagine that the RG bears a fairly strong degree of connectedness to the brave officer (BO). If psychological connectedness were what constituted numerical identity, we'd have the same problem on our hands that Locke did with his Memory Criterion, for it would have to imply that the RG both is and isn't identical to the AS: because the RG is connected to the BO, RG would have to be identical to BO, and because the BO is connected to the AS, BO would have to be identical to AS, and so by transitivity RG would have to be identical to AS, and yet because RG is in no way connected to AS, RG would also have to be *not* identical to AS.

We need, then, our account of numerical identity to appeal to overlapping chains of direct psychological connections to resolve this worry. But there is another worry to contend with. Suppose it were possible to transfer a memory trace of one of your experiences into my head. So perhaps I could wake up seeming to remember eating what you had for dinner last night. This would establish a minimal degree of connectedness between us. Would it establish any degree of *identity* between us, though? Clearly not: one connection does not identity make. What people emphasize in embracing a psychological theory of identity is that our identity across time is preserved only by there being an ongoing stream of *lots* of such connections. How many, though? This is a very difficult, if not impossible, question to answer. As Derek Parfit puts it, "[W]e cannot plausibly define precisely what counts as enough [connectedness]."[1] We seem to

1 Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), p. 206.

think that one connection isn't sufficient to contribute what matters to preserving identity, that 100% connectedness is definitely sufficient, and that some amount less than 100% is also sufficient (given that we forget things about ourselves and change in various other respects as well). But identifying that amount precisely is just beyond us. So what we may do instead is stay rather vague on this point: with respect to its contribution to identity, what counts is **strong connectedness**, where this will just refer to *whatever* amount of connections we would typically agree is sufficient.[1]

Strong connectedness couldn't be our criterion of numerical identity either, though, for the simple reason that it's not a transitive relation, which any proper criterion of identity must be (the identity relation just being the "=" relation). Strong connectedness, after all, is also a matter of degree. To return to our brave officer example, the RG could be strongly connected to the BO, and the BO could be strongly connected to the AS, and yet the RG might not be strongly connected to the AS. So for strong connectedness to play any significant role in the identity relation, we must appeal to overlapping chains of it. As a result, when Y is strongly connected to an intermediate stage who herself is strongly connected to X, call the relation one of *psychological continuity*.[2]

The Psychological Criterion of numerical identity is thus constituted (in part) by the relation of psychological continuity, which can preserve identity even between very distant, minimally connected or altogether unconnected, stages. This seems plausible, for many of us likely think that one's 80-year-old self is one and the same person as one's 10-year-old self, despite their not being very closely psychologically related. After all, there is still a single *stream of psychology* running from the 10-year-old to the 80-year-old, and it's that stream the Psychological Criterion points to as providing the identity relation.

---

1 In *Reasons and Persons*, Parfit suggests a more precise criterion: "[W]e can claim that there is enough connectedness if the number of direct connections, over any day, is *at least half* the number that hold, over every day, in the lives of nearly every actual person" (p. 206; emphasis in original). But this could at most be a guess.

2 This is just one way it might work, a case in which there's only one link in the chain separating X and Y. For some people, however, there may be many intermediate stages providing many links in the strong connectedness chain.

What about the "uniqueness" clause in the original formulation? It is meant simply to stipulate that there can be no other person-stage at t2, call him or her Z, who bears the same psychological relations to X that Y does. In other words, the relation between Y and X must hold uniquely in order for them to be the same person. Otherwise, for reasons discussed in the Divine Duplication case, Z and Y would have to be identical with each other, which they could not be. So the Psychological Criterion just eliminates this possibility by fiat: if there's ever a duplication that occurs, identity is lost.

Of course, this move is similar to the efforts of Dave Cohen when he presented Memory Criterion #3 to respond to the Divine Duplication problem, and you may very well think that Weirob's reply to it still applies here: this criterion would make my identity depend on the existence or non-existence of other people, which is absurd. Now on its face, this does seem rather crazy, but some advocates of the Psychological Criterion attempt to mitigate the craziness by making a very clever move to separate out personal identity from what *matters* in personal identity, so if it turns out that what matters in identity is preserved even in duplication cases, the fact that identity itself depends on the existence or non-existence of other people isn't as absurd as it might otherwise be. We will see how this move goes in the next chapter. For now, though, another response to the charge of absurdity might simply be that, while it may have this crazy implication, at least it doesn't yield any straight-out contradictions, as the Body Criterion does, and at least it provides a relation that's relevant and motivated, which neither the Soul Criterion nor the Brain-Based Memory Criterion are. So perhaps the absurdity involved is the least of four evils.

It's also worthwhile to emphasize the main advantage of the view, namely, it seems to account for rational anticipation and self-concern extremely well. If we continue with our assumption that personal identity is what grounds these prudential concerns and attitudes, then the Psychological Criterion makes a great deal of sense, for it looks as if I can rationally anticipate the experiences of, and have special concern for, only my psychological descendants. If some future person won't be connected to my current psychological stream, then it's hard to see how I could rationally anticipate his experiences

or have that special type of concern for his well-being.

There might be thought to be yet another advantage of the view: in one of its versions, it allows for the possibility of immortality. Following Parfit, we may distinguish between two versions of the criterion, depending on what one takes to be the appropriate *cause* of psychological continuity. On the **Narrow Psychological Criterion**, psychological continuity must be provided by its normal cause, namely, the persistence of the same brain, in order to preserve identity. Obviously, this version of the view couldn't allow for one to survive the death of one's body. But why think the proper cause of psychological continuity must be the same brain? Isn't it just the preservation of that psychological *stream* that matters? If so, and if that stream could be preserved via a different brain, or via any other method, that should suffice for what we want. This is the version known as the **Wide Psychological Criterion**, according to which psychological continuity provided uniquely by *any* cause is both necessary and sufficient for personal identity. And on this version of the view, one could indeed rationally anticipate the possibility of immortality, for as long as it is possible that God exists, cares, and constructs a person in heaven—but of course only *one* such person—who is psychologically continuous with you on earth, it is thus possible that *you* will survive the death of your body, and so it could be rational for you to look forward to doing so.

Yet another advantage of the view is that it helps to explain ordinary cases of *self*-identification, for how it is that you can typically know who you are when you wake up in the morning without having to check on the status of any substances such as your body or your soul. Instead, your relation to the person who got into your bed the previous night is (typically) evident, available to simple introspection, and it is clearly the sort of psychological relation we have detailed, consisting in memories, intentions-to-be-fulfilled, persistence of beliefs/desires/goals, and similarity of character. And it seems clear that if last night that person had anticipated your experiences this morning, he or she would have been rational to do so given your status as his or her psychological successor. Now there are exceptions, of course, cases in which people are just wrong about what memories or character they have (e.g., the poor deluded fellow who thinks

he's Napoleon), and it's possible that the psychological stream to which I have access doesn't hold *uniquely* between me and that person who got into my bed last night. Nevertheless, in ordinary cases our memories are genuine and we haven't been duplicated, so we can at least take it as the default position that what explains ordinary self-identification is the psychological continuity constituting the Psychological Criterion.

The basic thought behind the Psychological Criterion is rather simple: some past or future person cannot be me if that person is not *psychologically* related to me. This implies, then, that I cannot rationally anticipate the experiences of, or have reason for special concern for, those individuals to whom I won't bear that psychological relation. As mentioned earlier, this theory has probably had the most adherents among contemporary philosophers, at least until recently. It has declined in popularity over the past ten or fifteen years, however, because it seems to suffer from two general and serious problems: the ***Method of Cases Problem*** and the ***Essence Problem***.[1]

The *Method of Cases Problem* is fairly straightforward, challenging the method by which people are typically moved to accept the Psychological Criterion. Think, for example, about just some of the far-out cases we have envisioned: teleportation, the *Who is Julia?* brain/body transplant case, the brain rejuvenation scenario, Divine Duplication, and the possibility of waking up as a giant insect. All of these cases pump the general intuition that persons go where their psychologies go, so when we project ourselves into these imaginary scenarios, we're inclined to dismiss the importance of our bodies. As a result, if what matters to personal identity in these sorts of scenarios—where our bodies are prized apart from our psychological streams—is our psychology, then (the thought goes) some form of the Psychological Criterion must be true.

There are at least two problems with this method, though. First, how are we to take these cases up in our imagination? There are two ways, each of which is problematic. On the one hand, we could consider them as scenarios in which we, *as we currently are*—with our current values,

1 Several philosophers have discussed these problems, including David DeGrazia, Mark Johnston, Eric Olson, and Kathleen Wilkes.

beliefs, physical construction, and technological know-how—undergo their imagined procedures. But as we currently are, these scenarios are simply physically and technologically impossible! We have neither the knowledge nor the means to bring them about. So we are being asked to think about what *would* happen to us in scenarios that *could not* happen to us, given the way we are and the state of our current understanding. Obviously, then, if something could not happen to me, the question of what *would* happen if it could is a nonstarter. On the other hand, we could consider these scenarios as occurring to creatures for whom they aren't physically and technologically impossible. But such creatures would be very different from us, so different it would be difficult to imagine what their lives could possibly be like. Yet on this way of understanding the method of cases, we would be asked to explain what would happen to *us* in scenarios in which *we would not exist*. Instead, it would be these other sorts of creatures that exist, and how could we have anything enlightening to say about what would happen to them? Thus we have a real problem: the scenarios would be possible only for creatures unlike us, in which case we couldn't draw any stable or illuminating philosophical conclusions from our imaginative consideration of them.

One possible response to this worry, however, is that there seems no reason in principle to think that the creatures who would exist in the technologically-advanced future couldn't be exactly like us in terms of our *values* and *conceptual knowledge*, and that's the similarity that matters for yielding viable results from these thought experiments. In other words, why can't we envision people psychologically just like us who simply happen to have a technology we don't currently have (perhaps we can imagine that aliens have just landed and given us the gift of teleportation)? If so, then why couldn't we learn something from these sorts of thought experiments? Surely we can put ourselves into the shoes of such creatures sufficiently to figure out what we would believe and how we would feel if these things were to happen. And that's all that's required from the method in question.

A second problem, however, is much more insidious: the Method of Cases, goes the charge, can actually yield contradictory intuitions. The objection here is based on a pair of cases first articulated by Bernard

Williams. In the first case, a person with what we'll call Body A will have his entire psychology downloaded to a computer, which will temporarily erase all the contents of his (Body A's) brain. Simultaneously, another person (with Body B) will have his entire psychology downloaded to the same computer, temporarily erasing all the contents of his (Body B's) brain. Then the first set of psychological contents will be uploaded into Body B's brain, and the second set of psychological contents will be uploaded into Body A's brain. What happens to each of the original people? This is, of course, just another version of our now-familiar "body swapping" thought experiments, and it's again one in which it seems clear that in each case the original person got a new body, going where his psychology went. So the person who originally had Body A now has Body B, and the person who originally had Body B now has Body A. Indeed, suppose you were the original person with Body B, and I were the scientist about to do the procedure, and I told you that after the downloading and uploading, I would give $1 million to the person who winds up with Body A and I would torture the person with Body B. You would likely be quite happy about this prospect, which indicates you think that you'd be the one with Body A after the procedure. So far so good.

But now consider a different case. Suppose you're kidnapped and a scientist tells you that he's going to torture you. "But first," he tells you, wild-eyed, "I'm going to erase your memories, beliefs, desires, intentions, and all the rest of your specific psychological characteristics!" "Oh great," you think, "I'll lose my mind, and then I'll be tortured." "But I'm not finished!" exclaims the scientist. "After I delete all your psychological characteristics, I'm going to implant in you all the psychological characteristics of your neighbor, constantly-stoned Fred!" "Wonderful," you now think, "I'll lose my mind, then I'll be deluded, and then I'll be tortured. What a lovely day I have in store." Then suppose the scientist tells you that, while this is going on, he'll be implanting a copy of your psychological characteristics into the brain of your neighbor Fred, and then he'll give Fred $1 million. This will not likely cheer you up. Notice, then, that you believe in such a case that *you* will persist through all these changes. But now the problem should be obvious, for all of this is just a different way to describe the exact same case from above,

with you in the position of the Body B person. In that first case, though, your intuitions were likely that identity is preserved entirely by psychological continuity, whereas in this second case, your intuitions are likely that identity is preserved entirely by *physical* continuity. But now we've got seemingly contradictory intuitions *on precisely the same case*. A mere difference in description of a case shouldn't yield contradictory identity-judgments, and yet it does, which should lead us to have serious doubts about the viability of intuitions pumped by the Method of Cases generally.

Various replies have been given to this puzzle over the years, some of them fairly complicated. Given our purposes here, however, we might simply think about two questions. First, what should we think of the last inference given, that the Method of Cases *generally* is in trouble, given the troublesome pair of cases articulated by Williams? On its face, this seems quite a hasty generalization. Perhaps, after all, the Williams example is the only thought experiment, or one of the only few, that yields the contradictory intuitions. Without some more examples of the problem, then, we might still be justified in deploying the method to yield evidence from our intuitions in other cases like teleportation, Divine Duplication, and the *Metamorphosis*-style examples. And these cases still seem to produce intuitions in line with the Psychological Criterion.

A second question to consider, though, is this: how much support does the Method of Cases actually provide to the Psychological Criterion? Suppose, for example, that we take the Williams objection to be decisive against the Method of Cases. Does that mean we should *therefore* abandon the Psychological Criterion? Clearly not, for there are certainly other considerations in its favor. One is its facility in dealing with the self-identification phenomenon. Another is its facility in accounting for rational anticipation generally. And yet another might be its facility in handling certain real life cases, such as that of the Hensel twins. So we shouldn't think that, even if the Method of Cases is to be abandoned (and we have just seen reason to doubt the motivation for such a response), the Psychological Criterion is undermined as a result. One might think that one of its pillars of support has been lost, but that would not prevent the remaining pillars from providing nearly as much support as before.

The Essence Problem, on the other hand, discussed briefly in the introduction, is far more troublesome. It is also rather complex. It begins, however, with a seemingly simple question: what am I? This is, of course, a question about *membership in a kind*: to what *kind* do I belong? Now as it turns out, you and I (and all other readers of this book) are *many* kinds of things. Speaking for myself, I am an adult, a professor, an author, a husband, a stepfather, a driver, a voter, a homeowner, and many other things. More generally, I am a human being, an embodied mind, a biological organism, and a person (as are you all). But which of these many kinds in which we're all members is most fundamental? Is there some kind of thing that, if we weren't *that*, we wouldn't exist at all? This is to ask the question of essence: what am I *essentially*? What we are looking for may be called a **basic kind**. And there's a truism among metaphysicians that a determination of the basic kind to which a thing belongs—determining the essence of that thing—yields the necessary *identity conditions* of that thing as well. After all, if some object O has some essential property X, a property without which it couldn't exist, then in order for O to continue to be the same object over time, it must continue to have X.

Here's an illustration of the dependence of identity conditions on a determination of what kind a thing belongs to. Imagine that there's a statue in the park made of a big lump of bronze; after a while, people get tired of it, and the statue gets melted down and the melted-down lump is dumped in a warehouse. Now suppose somebody asks you: "That thing that used to be here in the park—does it still exist?" You might answer, "No, that statue was melted down and doesn't exist any more." Or you might answer, "Yes, that lump of bronze still exists, now gathering dust in a warehouse." Whether the thing that was in the park is identical with the thing now in the warehouse depends on what kind of thing the thing in the park was most fundamentally. If it was fundamentally a statue, then it couldn't be identical to the thing in the warehouse, for it would no longer exist; if it was fundamentally a lump of bronze, then it would be identical to the thing in the warehouse, for its very essence has survived. So which is the *right* way to think of the park-object? That depends on

what its essence really is, that is, it depends on what we couldn't conceive it to exist without. Once we figure this out, we'll know the basic kind of the park-object.

So what is my basic kind? What am I essentially? I can very clearly conceive myself as not being an adult, professor, husband, stepfather, driver, voter, and homeowner. Indeed, when I was a young teenager, I was none of these things. So none of these is my basic kind. And the same goes for the rest of you. Perhaps then our basic kind is one of the more general kinds mentioned earlier. But which one?[1]

Advocates of the Psychological Criterion seem to suggest an answer: what I am essentially is a person, and persons are, by definition, psychological beings, which means that my identity across time must necessarily involve the persistence or continuity of my psychology.[2] Here advocates of the Psychological Criterion follow John Locke, who famously defined a person as "a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places; which it does only by that consciousness which is inseparable from thinking, and, as it seems to me, essential to it...."[3]

But is this true? There are actually several serious problems with thinking of ourselves as being essentially persons:

1. *The Fetus Problem*. If what I am essentially is a person, a psychological being, then how can we make sense of the following common thought: "I was once a fetus"? Surely this is coherent. For instance, suppose your

---

1 Incidentally, for the sake of argument here, we are sharing in the assumption of most philosophers working on personal identity today that every concrete object that exists belongs most fundamentally to one and only one kind, a kind which provides the object's identity conditions across time. This isn't a universal assumption, though, and there are alternatives to it one should keep in mind. One alternative is the view that certain objects simply don't have a fundamental essence. Another alternative is the view that an object's essence doesn't provide its identity conditions, or at least the identity conditions that matter. Keep these alternatives in mind as we proceed.

2 A *person* is being thought of here as necessarily having a psychology, and not merely as a human organism, contrasted with other organisms. In this rather special technical use of the word, a live human organism lacking all higher brain functions and permanently unconscious, thus without any psychology, wouldn't be a *person*.

3 John Locke, "Of Identity and Diversity," in John Perry, ed., *Personal Identity* (Berkeley, CA: University of California Press, 1975), p. 39.

mother still has the photo from her sonogram, and as you see it, you say, "Wow, that was me?" To take another instance, suppose you had fetal alcohol syndrome, having certain psychological difficulties as a result of your mother's drinking while she was pregnant. Wouldn't it be correct to say that *you* had been damaged while in the womb, and that perhaps you are now owed some sort of compensation because you were harmed as a fetus? Nevertheless, if you are essentially a psychological being, and a fetus (especially prior to developing a brain) is not, then you simply could not have been a fetus. But this seems incorrect, given our natural and common way of talking. Just as it is clear that I existed prior to being an adult, so too it seems clear that I existed prior to being a person.

2. *The PVS Problem*. Suppose that you get into a terrible accident that destroys your brain's capacity for consciousness, while leaving the brain stem intact, resulting in your body's being in a permanent vegetative state (PVS) in the hospital, permanently unconscious but with the capacity for spontaneous breathing, a heartbeat, and other biological functioning. Isn't the correct description that *you* are the one in the PVS? Certainly this is what your devastated parents, spouse, and friends would think as they continued to visit that hospital room. Nevertheless, there seems no way to render this way of thinking coherent if you are essentially a person, for you could not be this individual in the PVS, given its lack of psychology: if the individual in the PVS isn't a psychological being, then this individual isn't a person, and if this individual isn't a person, then this individual can't be you. But again, this seems to be the wrong answer. Just as it is clear that I will continue to exist after retiring as a professor, so too it seems clear that I will continue to exist after going into a PVS.
3. *The Person/Animal Problem*. Suppose I'm essentially a person, a psychological being. When did I come into existence? Presumably, I appeared when my psychological motor started running, likely at the late fetus/early infant stage. What, then, happened to that pre-psychological organism, that human animal? Did it die? If so, what happened to its remains? Can there be death without any remains? Perhaps, then, that organism just disappeared. Nevertheless, this is quite unlikely; organisms don't just disappear, as far as we know. Perhaps, then, that organism (the human animal) still exists, its existence somehow overlapping with mine (the person). But then when

looking at my body you would see two distinct beings, the human animal and the person. Indeed, these would be two wholly distinct *substances*. This would be, at the least, quite odd.

A similar problem attaches to the other end of the life spectrum. If I am a person, when do I go *out of* existence? It would seem that when my consciousness ends, so do I. But then does the PVS patient *begin* to exist at my (the person's) death? This seems unlikely. After all, that patient is a biological organism, and biological organisms are typically brought into existence by some kind of birthing process, whereas there just is no such process here. Did it then simply *appear*? Again, that is quite unlikely: organisms neither simply appear nor disappear. Perhaps, then, it was in existence from the fetal stage, and thus overlapped with me over the course of my life. But then, once more, when looking at my body now you'd be seeing two distinct substances, the animal and the person, which would be very strange.

These are just some of the problems that have motivated the development of a very different thesis about our essence, and thus a very different thesis about our identity across time.

The Biological Criterion

The alternative view is that we are essentially *human animals*, specific types of biological organisms. While we are persons during much of our lives, we are not *essentially* persons. Instead, personhood is just one stage we animals live through, one in a series of non-basic kinds to which we temporarily belong. Other non-basic kinds include fetus, infant, toddler, adolescent, teenager, adult, senior citizen, and, perhaps the kinds of the senile, the demented, or the PVS. Further, if we are essentially animals—biological creatures—then this fact yields its own criterion of personal identity:

> **The Biological Criterion**: *If X is a person at t1, and Y exists at any other time, then* X=Y *if and only if Y's biological organism is continuous with X's biological organism.*

There are some important features of this criterion worth discussing. First, you will no doubt have noticed a key difference between this formulation and our previous criteria, namely, this one is broader. In all of our previous formulations, the criterion of personal identity told us what makes X and Y the same *person*. The current criterion, however, purports to tell us what makes something that is a person at one time (X) the same *thing* as a Y that may or may not be a person at a different time. This is simply because advocates of the Biological Criterion believe that what I am essentially was not always, and may not always remain, a person, so in order to capture my identity conditions across time, we should not restrict the class of those things to which I might be identical only to the class of persons. I could, after all, be one and the same thing as my future PVS stage. Of course, no one likes to think of himself or herself as a *thing*, so many advocates of the Biological Criterion use the term *individual* instead, rendering an alternative formulation of the criterion as follows: *X (a person) at t1 is the same individual as Y at any other time if and only if Y's biological organism is continuous with X's biological organism*. I will use these two formulations interchangeably throughout the book.

The second thing to notice is that this criterion is tracking the *continuity* of organisms. What this means is fairly simple: one organism is continuous with another just in case the life-sustaining functions of the former organism are inherited by—they continue on in—the latter organism. Another way to think of the continuity involved here is that it just describes the relation that obtains when a biological organism is, in principle, uniquely traceable across space-time.

Third, while it may seem as if the Biological Criterion is just a version of the Body Criterion with more syllables, there is actually a subtle but important difference between the two. As Eric Olson has explained, while it seems perfectly clear what a biological organism is, it is actually surprisingly unclear just what a human body is. Indeed, what exactly is your body? There are two possibilities: either (a) your body is the material object that you can feel in some direct way and are able to move just by willing it, or (b) your body is just whatever it is that we're talking about

when we attribute certain physical, spatial, or time-based (temporal) properties to you in our ordinary ways of speaking.

On the first possibility, whatever it is you can feel and move directly is your body. But there are all sorts of problems with this account. For one thing, those who are paralyzed cannot feel or move what are still certainly parts of their bodies. Relatedly, I cannot move various internal organs, like my liver, at will. Does that mean that my liver is not part of my body? In a different vein, I can move my right foot directly. Does that mean my right foot is my body? Or suppose I'm holding a pencil in my hand. When I write something down, don't I move the pencil directly? Does that mean it's my body, or perhaps just part of my body? Finally, what do we say about those with prosthetic limbs? Are those limbs included under the concept of "my body"?

On the second possibility, we carve out the concept of your body in terms of ordinary ascriptions of physical properties to you. So when we say, "You're tall," we're ascribing the property of tallness to you, and insofar as that's a physical property, whatever it is that bears that property *just is* your body. Similarly, when we ascribe certain spatio-temporal properties to you, for example, "I saw you at the gun show playing with the bazookas yesterday," then whatever it is that took up space at that particular time and place just is your body. Ultimately, then, whatever has *all* the physical and spatio-temporal traits we could ordinarily attribute to you is what counts as your body. The problem here is that this definition assumes exactly what is in question here—that I am my body—because it renders all physical and spatio-temporal properties attributable to *me* properties of my body. But there are some people who disagree, who think that I am distinct from my body. If they are right, then there would actually be *two* bearers of the attributed physical and spatio-temporal properties: my body and me. By insisting there's only one such bearer, however, this second attempt to provide a clear conception of "human body" also assumes what's at issue, and so fails as well.

What all this means, therefore, is that the concept of a "human body" is simply too unclear to do any real work for us as part of a criterion of personal identity. As it turns out, then, Weirob was relying on an

unusable concept of who (or what) she was. The advocate of the Biological Criterion, however, has no such worries. This criterion provides meaningful persistence conditions for me across time in virtue of my being an *animal*, not a human body, and what counts for our purposes as an animal, as a living human organism, is easily conceptualized as falling under a commonsense biological category. (Although we will see in the chapters to come that the boundaries of what's included under the concept of a human organism are less clear-cut than the advocates of the Biological Criterion would have us believe.)

So much for the details of the Biological Criterion itself. Why should we believe it? There are several considerations in its favor, some of which we have already run across. We will sometimes put these in terms of its relation to its chief rival, the Psychological Criterion:

1. *The Biological Criterion seems to provide a more plausible story about our essence than the Psychological Criterion.* The Psychological Criterion, remember, seems to imply that we are essentially persons, but if that's the case then it's very difficult to make sense of perfectly ordinary ways of talking, like "I was a fetus," and "If I go into a PVS...," and it's also difficult to make sense more generally of the relation between persons and their animal organisms. The Biological Criterion, however, easily handles these worries, for it identifies us as essentially animals, in which case I—this individual that is now in its "person-phase"—was indeed a fetus, could eventually be in a PVS, and my animal organism and I are simply one and the same thing.
2. *The Biological Criterion allows there to be a tight and direct connection between the metaphysical criterion of identity and the epistemological criterion of identity (perhaps even more so than the Psychological Criterion).* Recall the distinction between these two types of criteria from the Introduction. A metaphysical criterion of identity will tell us what makes X and Y identical. An epistemological criterion of identity will tell us how we can identify whether or not X and Y are identical. Many people think, then, that it would be good to have a metaphysical criterion of identity that would make it easy to identify when that criterion has been met in the real world, and the

Biological Criterion seems tailor-made to provide just this. After all, how is it that we typically reidentify others, identifying whether or not the person we're dealing with now is the same person as the one we dealt with earlier? By recognizing their human organisms. True, we sometimes reidentify people without seeing or hearing them (via e-mail, say), but here it might be thought that what we're doing is reidentifying their organisms *indirectly*. The Psychological Criterion, on the other hand, might be thought to have more difficulty in this arena, for we cannot reidentify streams of psychology directly at all (I can't somehow *see* your psychology), nor does it seem as if we are even doing so indirectly sometimes: when I see your face across a crowded place, I know it's you, without making any further inferences about your psychology. In this respect, at least, the Psychological Criterion may be as irrelevant as the Soul Criterion.

3. *The Biological Criterion is broader and more inclusive than the Psychological Criterion, providing persistence conditions for human animals that are not, or won't be, persons.* Suppose an anencephalic[1] infant is born without a cerebrum, and this infant manages to live for a month. Surely the month-old anencephalic infant is the same individual as that just-born anencephalic infant, even though neither possesses (or will ever possess) the capacity for consciousness. The Psychological Criterion must thus remain silent about the persistence conditions for these human infants, whereas the Biological Criterion includes them as *one of us*, human animals, whose persistence conditions are the same for members of that group with or without a psychology. Insofar as we are inclined to think that, at least with respect to identity, the cases of humans with and without the capacity for consciousness should be treated alike, the Biological Criterion has a distinct advantage over the Psychological Criterion.

There look to be some real advantages to the Biological Criterion. However, recall that the Psychological Criterion had its own set of advantages as well, so in order to engage in a fair comparison of the two views,

1 *Anencephaly* is lack of a major portion of the brain, skull, and scalp, resulting from improper fetal development. The lack of a forebrain means that the child will be permanently without any conscious functions.

we need to have before us some of the main *problems* associated with the Biological Criterion.

1. *The Conjoined Twins Case.* One of the real problems for the Body Criterion, recall, was the case of the Hensel twins, who seem to have one body but are clearly two persons. If what makes X and Y the same person is their having the same body, and Brittany and Abigail have the same body, then they would have to be the same person, according to the Body Criterion, which is obviously false. Wouldn't the same be true of the Biological Criterion, however? Wouldn't Brittany and Abigail be the same human organism, which would imply that they are identical with each other? Not necessarily. David DeGrazia endorses the possibility that theirs could be a rare case of two *overlapping* organisms. After all, they (mostly) have two distinct sets of organs above the waist, that is, they each have their own hearts, brains, stomachs, and so forth. And insofar as these organs are what typically provide the regulatory and sustaining aspects of living organisms, Brittany and Abigail can easily be thought of as two organisms that overlap to some extent.

This is too quick, however, for we might just as easily point to other features of the Hensel twins that strongly suggest that they are *one* organism. For example, they have a single skin, a single liver, a single urinary tract, a single blood stream, a single immune system, and a single reproductive system. Furthermore, even their distinct sets of organs function together in the integrated way distinctive of living organisms, such that if one sister's set of distinct organs were to fail, the other sister's organs would also fail immediately thereafter. But if we think of the death of organisms as consisting in the irreversible cessation of the integrated functioning of its organs, then there would be just one death here, the death of a single organism.

Obviously, the answer here depends on how we define "organism," and this is a matter of some controversy. Now there are some cases of conjoined twins where DeGrazia's interpretation is clearly correct. The original "Siamese" twins, Eng and Chang Bunker, were joined at the chest by a five-inch-wide band of flesh (and their livers, while individually complete, were also fused). In such a case, the thought that they were distinct organisms

that very slightly overlapped is a natural one. But suppose there were a case at the opposite end of the conjoined spectrum, one in which there were two heads on one body, but even the heads were partially fused, having one brain stem, say, but having two distinct faces—two eyes, noses, and mouths—and two distinct centers and streams of consciousness (given distinct cerebrums). The interpretation of two distinct but overlapping organisms would become much harder to maintain in such a case. Individuating organisms by pointing to distinct cardiopulmonary regulatory systems, say, something that could work to render the Hensel twins distinct organisms, wouldn't work in this case, given that there would be only one heart and one set of lungs. And individuating the organisms in virtue of their autonomic control centers wouldn't work here either, given that these twins would share a single brain-stem. Indeed, it's difficult to think of any non-arbitrary way to individuate these (hypothetical) twins as distinct organisms.

When faced with such a case, DeGrazia (a defender of the Biological Criterion) holds out the possibility that this could be a case akin to Multiple Personality Disorder (MPD), in which there is indeed just one organism, but one with two distinct centers of consciousness. This seems too much of a stretch, however. For one thing, the centers of consciousness could be simultaneously engaged and each could be continually aware of the other, which is not the case for most of those with MPD. But aside from the analogy to MPD, DeGrazia's view would imply that here we would *not* have two individuals—he still maintains the one-to-one correlation between organisms and individuals—while in the Hensel twins case we *would* have two individuals (overlapping organisms). But surely what leads us to believe the Hensel twins are both distinct persons *and* distinct individuals—their communication, their disagreements, their conscious coordination, their insistence on individuality, their independent ways of thinking—all of these features could be present as well in the more extreme case. It is hard to believe, then, that if the Hensel twins would be distinct individuals, our imagined extreme conjoined twins would not be as well.

2. *The Corpse Problem*. When I die, I will leave behind a corpse. But what is that corpse's relation to me, the human animal? Upon my death,

there will still be a physical continuity between me and my dead body, but doesn't the Biological Criterion then imply that I will *be* that dead body, that it will be the same individual as me? What is the advocate of the view to say here?

There are generally three replies one might give. First, one could embrace the implication, and affirm that I will indeed be that corpse, that that's what *I* will be at some point in the future. But although some writers have embraced this implication, it seems wildly implausible: surely that corpse will not be *me*. Indeed, our ordinary practices strongly support the intuition that, upon our deaths, *we are no more*. Thus the grieving and mourning that takes place when our loved ones die. If we thought they still existed among us, such behavior would be odd, if not downright incoherent. To remain plausible, then, the Biological Criterion must accept that I go out of existence with the death of my biological organism. But if I am not my corpse, then it must be a numerically distinct object from me.

This fact leaves the advocate of the Biological Criterion with two ways of dealing with the issue: (a) when I—the human animal—cease to exist at death, my corpse—a distinct individual object—pops into existence; or (b) my corpse-to-be, a distinct individual object, has existed all along, coinciding in space-time with my living biological organism. The latter option is independently quite implausible, but even worse, it undermines one of the main motivations for the Biological Criterion in the first place, namely, to avoid the problematic implication of the Psychological Criterion that persons and human animals are both numerically distinct objects but also both wholly coincide. If option (b) were taken, though, the Biological Criterion would be in precisely the same jam, having to make sense of the bizarre fact that seated in my chair at this moment are two numerically distinct objects, me (a human animal) and my corpse-to-be (which would suddenly make things very creepy).

To avoid this implication, then, the best bet for the advocate of the Biological Criterion is to opt for (a), that when I cease to exist, my corpse then pops into existence. And at first glance, it does seem we are talking about very different sorts of objects: the animal I am is alive, its various

organs are functionally integrated, it uses resources from its environment to maintain a stable regularity, thus preserving its form over time, and so on. My corpse, on the other hand, will have non-functioning organs, will make use of no environmental energy for self-sustenance, and will eventually lose its form more or less entirely over time. So why not think that the corpse comes into existence upon my exit?

One reason to be hesitant about such an answer, though, comes from the worry that we may not have a firm grip on what constitutes the death of an organism, and this uncertainty will carry over into uncertainty about *when* the corpse comes into existence. It would be rather odd, though, that we would have such trouble marking the difference between two such categorically, qualitatively, and numerically distinct objects. A further worry is that the definition of death might wind up being a matter of pure stipulation. But surely a matter of metaphysical reality—the coming-into-existence and going-out-of-existence of numerically distinct objects—couldn't depend on convention in this way.

3. *The Transplant Intuition*. Regardless of the considerations in favor of the Biological Criterion, our intuition in the *Who is Julia?* case, that Julia is the survivor in Mary Frances's old body, likely remains strong. This intuition cuts sharply against the Biological Criterion, however, for that criterion maintains, along with Weirob, that as long as Mary Frances's regulatory biological mechanisms remain in place in her original body the individual that is Mary Frances remains as well, even if that organism gets a new brain (or cerebrum, which is all that's needed to make the point). So the survivor is a deluded Mary Frances, someone who *thinks* she's Julia, but is sadly mistaken.

This implication is unlikely to sit well with many of us, though. After all, we typically identify with our psychologies, thinking that we are essentially psychological creatures, and if our brains underlie that psychology, then we go where our brains (or cerebrums) go. It is difficult to believe, then, that if our cerebrum were removed, and replaced with someone else's, that *we* would somehow remain, and remain permanently deluded. That individual, after all, would have no psychological connection to us whatsoever, and this worry leads us to our final problem with the Biological Criterion.

4. *The Prudential Concerns Problem*. Remember how we started all of this off: we wanted to know what the rational grounds for anticipation or special concern are. We have been assuming that personal identity is at the very least a necessary condition of rational anticipation and special concern: I can't rationally anticipate some future person's experiences or have that special sort of *self*-concern unless he will be me. But we might plausibly think that what we wanted to know assumed something even stronger, namely, that identity is a *sufficient* condition of both rational anticipation and self-concern, that is, my identity with some future person is what in fact provides me with sufficient reason both to anticipate his experiences and have special concern for his well-being.

If we go with this stronger assumption, though, the Biological Criterion fails, for it can't be solely in virtue of the fact that he is my biological continuer that I have a reason to anticipate, say, some future person's experiences. To see why, simply consider the case in which I fall into a PVS. Would I have any reason whatsoever—let alone a sufficient reason—to anticipate this biological continuer's experiences? Surely not, for the simple reason that he will be incapable of undergoing any experiences for me to anticipate! Similarly, we might hold that I have no reason to have any sort of special concern for the well-being of my PVS descendant, given that, because he would lack the capacity for conscious experiences, he would lack the capacity for well-being as well. The stronger sufficiency assumption, then, favors the Psychological Criterion, for it maintains that any future person who is me will at least be my psychological descendant, and so will at least be a conscious experiencer.

One reply here would simply be to deny the sufficiency assumption. Perhaps it's a mistake to assume that identity is sufficient for making anticipation rational. After all, it may not be rational for me to anticipate the experiences of my 90-year-old self (assuming I live that long!), insofar as he's likely to be *very* different psychologically from me, despite the fact that he'll still be psychologically continuous with me. Nevertheless, we can capture what seems important about the sufficiency assumption, without denying this claim about my 90-year-old self, by making a crucial distinction between what's rationally *required* and what's rationally *permissible*.

Surely one is not rationally required to anticipate the experiences of one's 90-year-old self; indeed, one may not be rationally required to anticipate the experiences of *any* of one's futures selves. But certainly it is rationally *permissible* to do so, that is, it is *not irrational* to do so. The sufficiency assumption, then, could simply be the claim that personal identity is what makes anticipation and special concern rationally permissible. And if this is the case, then the Biological Criterion still fails the test, for it can't be solely in virtue of some future individual's being biologically continuous with me that it is suddenly rationally permissible for me to anticipate his experiences. Something more is needed, and that something more must be psychological in nature.

Nevertheless, even if we were to deny the sufficiency assumption, we've still got the necessity assumption to deal with, namely, the claim that personal identity is *necessary* for rational anticipation and special concern. And with respect to even this assumption the Biological Criterion comes up short, in light of the transplant cases. Suppose Julia knew her cerebrum were going to be transplanted into Mary Frances's body, and that the resulting person would be exactly similar to Julia psychologically. Many of us would think it would be perfectly rational for Julia to anticipate the experiences of, and have special concern for, the survivor. If we persist in assuming personal identity is necessary for that activity, though, only the Psychological Criterion passes this test; the Biological Criterion has to maintain that the survivor is Mary Frances, and so it would not be rational for Julia to anticipate *anyone's* experiences, say, for she would be dead. But again, this will seem wrongheaded to many of us.

Summary

We are left, then, with several reasons for and against both of the main theories of personal identity, and these reasons are summarized in the following chart:

	THE PSYCHOLOGICAL CRITERION	THE BIOLOGICAL CRITERION
CONSIDERATIONS IN FAVOR	• Does well in "intuition pump" science fiction cases • Accounts for self-identification very well • Explains the rationality of anticipation	• Incorporates the most plausible account of our essence • Accounts for third-person reidentification very well • Includes a plausible story about the identity conditions of non-person humans (e.g., anencephalic infants, fetuses, PVS patients)
CONSIDERATIONS AGAINST	• The Method of Cases problem • The Essence Problem (which includes the Fetus Problem, the PVS Problem, and the Person/Animal Problem)	• The Conjoined Twins case • The Corpse Problem • Can't account very well for the Transplant Intuition • Can't account very well for rational anticipation

So which side wins? This is obviously not an easy call, for there are powerful considerations both in favor of, and against, each theory. And we must not make the mistake of thinking that, for example, because the Biological Criterion has more bullet points against it than the Psychological Criterion does, it is somehow worse off, for while these are indeed real problems for the theory, the Essence Problem is a far more serious, or weighty, worry for the Psychological Criterion than any of these. And while it may also seem as if its inability to account well for rational anticipation counts as a devastating blow to the Biological Criterion, we cannot forget that one might be interested in the issue of personal identity *independently* from its relation to our practical concerns—one might think of it solely as an interesting puzzle in metaphysics—and so from that perspective this so-called problem may be no problem at all (more on this point in the final chapter). But at any rate, as things stand it actually seems as if the *sets* of problems for each theory are roughly equal in seriousness.

It is quite unclear, then, just how one side might convince the other to join its ranks. But if we cannot determine which theory of personal

identity is correct, how can we determine the right answer to our identity-related ethical questions? After all, if we apply the Psychological Criterion to the problem of abortion, say, we are likely to get a very different answer from what we'd get if we applied the Biological Criterion. So what shall we do?

There are three general options. First, we might devote much more energy than we have to explore possible defenses against the objections raised against one of these theories. This is the work that many advocates of each theory have recently undertaken. The idea is to show how the objections raised against the *other* side's theory are insurmountable, while the objections raised against *one's own* theory are, well, surmountable. Because both sides have extremely smart advocates, though, we might be warranted in a persisting skepticism that the standoff will end via this method anytime soon.

A second option is to gain additional data about the viability of each theory by seeing how plausible its implications are for all of our practical concerns, both prudential and moral. Up until now, we have been considering how the views account for only our prudential concerns, but it may be that once we understand their implications for our moral concerns—concerns having to do with abortion, advanced directives, moral responsibility, compensation, and so forth—we will come to see that one theory is clearly superior to the other (at least in the way it accounts for such concerns). This is the strategy we will employ in Part B of the book, in fact. Of course, there are some genuine problems with this approach as well—not the least of which is that the correct criterion of personal identity may not answer to our practical concerns at all!—but we will save discussion of these concerns for the final chapter.

A third option is to explore an entirely different path, to find a new alternative to both theories. This is to recognize the standoff and, in a way, to try to move beyond it, to show that there is still something important to say about how identity relates to our practical concerns that simply doesn't depend on the standard criteria. This is a rather radical approach to the issue, of course, but it may be the best way in which to proceed in light of our current standoff. At the very least, we need to consider whether or not

it is possible to find such an alternative. This, then, is what we will attempt to do in our next chapter.

WORKS CITED OR REFERENCED IN THIS CHAPTER

DeGrazia, David. *Human Identity and Bioethics*. Cambridge: Cambridge University Press, 2005.

Johnston, Mark. "Human Beings." *Journal of Philosophy* 84 (1987): 59-83.

——. "Reasons and Reductionism." *The Philosophical Review* 101 (1992): 589-618.

Locke, John. "Of Identity and Diversity." In *Personal Identity*, ed. John Perry. Berkeley, CA: University of California Press, 1975, pp. 33-52.

Olson, Eric T. *The Human Animal*. Oxford: Oxford University Press, 1997.

Wilkes, Kathleen. *Real People*. Oxford: Oxford University Press, 1988.

Williams, Bernard. "The Self and the Future." In *Problems of the Self*. Cambridge: Cambridge University Press, 1973, pp. 46-63.

CHAPTER THREE

Alternative Approaches

In this chapter, we consider two fairly radical alternatives to the standard approach to articulating the relation between personal identity and our practical concerns explored in the last chapter. There we were left with a kind of standoff: both the Biological Criterion and the Psychological Criterion have serious advantages and serious disadvantages, and it's hard to know which one is more plausible (or if *either* is all that plausible) as a result. In light of this sort of standoff, various authors have been motivated to propose intriguing new possibilities for understanding the relation between identity and ethics. To this point, we have been assuming, along with the advocates of the standard approaches, that what matters to our practical concerns is some criterion of numerical identity. The first alternative we will discuss in this chapter, however, denies that *numerical* identity is what matters to our practical concerns, whereas the second alternative we will discuss denies that *identity* is what matters at all.

Narrative Identity

To this point we have been trying to come up with a workable criterion of numerical identity, an account of what makes a person at one time identical to some person or individual at some other time. This is because,

quite simply, many advocates of the standard approaches have assumed that numerical identity is the only sort of identity relevant to our practical concerns. As it turns out, though, there is another sort of "identity" that may be what's actually important here, a more everyday sense of the term familiar from cases in which someone undergoes an "identity crisis." More generally, this alternative sort of identity has to do with what makes us *who we really are*. Marya Schechtman has been the most articulate in developing what has often been an unclearly-presented position, so in laying out the view we will (mostly) follow in her footsteps.

To understand the sense of identity in question, consider a few cases. Suppose Kyle has been out of college for a few years. He was an English major, but though he enjoyed it, he never considered going on to graduate school to study more of it. After graduation, he returned home to live with his parents, and he has since bounced around from low-paying job to low-paying job. He parties on the weekends, sleeps in late during the day, goes to his job (when he has one), and plays a little music occasionally on his guitar. His parents have grown very frustrated with him, and Kyle is feeling the pressure to "do something" with his life, but he just doesn't know what that "something" is or should be.

Consider next Jack, a cop for ten years. Over the past year, sparked by a new love interest, he's been studying Buddhism, and he has come to consider himself to be a fledgling Buddhist. More and more, though, he sees a conflict between his job and his new religious beliefs. Being a police officer may require him to shoot someone in the line of duty, whereas his Buddhism requires him to be a pacifist, never to react with violence to the deeds of others. He is growing more and more concerned over this conflict (and other conflicts, including those about the various attitudes he should take to other people), and he is coming to the realization that only one of these lives can be lived honestly and wholeheartedly. So which is he really, he wonders, a cop or a Buddhist?

Consider finally Sarah, for many years a miserable and pitiful person. She was an alcoholic and a misanthrope, getting fired repeatedly for her hateful comments or her absenteeism, spiraling ever deeper into debt and, for a little while, homelessness. One day she hit rock bottom,

finding herself broke and alone in an alley, mysterious bruises on her arms and face, and a crushing hangover. "That's it," she thinks, "This can't be who I am." She contacts a relative, who agrees to take her in on the condition that she embark on a twelve-step program, which Sarah is more than willing to do. She thus starts a painful program of recovery, at the same time trying to work on her social skills, with the intention of being sober and the owner of a new life in ten years. After much hard work, she finds herself ten years down the line as a sober, industrious, and well-liked person. In a quiet moment one day she reflects back on her former life and thinks to herself with a kind of wonder, "Wow, I really did it!"

What we have seen here are three different arenas in which a non-numerical sense of identity is in play. Kyle simply doesn't know who he is. He feels, in a way, unformed as a person, without any real identity, and there simply seems to be no obvious direction for him to go to find one. Jack, on the other hand, also doesn't know who he is, but his bafflement isn't due to having no direction; instead, it's due to having too many directions. His commitments are in tension, pulling him down two different and exclusive paths, and as he stands at the point of their divergence, he's uncertain which way he'll go. Sarah, finally, was taking one horrific path with her life but managed to make the radical decision to leave it for another. She was miserable being who she was, and she finally became determined that she was not going to be the type of person living that kind of life, and so she embarked on a series of changes that would make her into the type of person she could eventually be happy being.

These scenarios bring out the sense of identity at issue, which responds to what is known as the *characterization* question. To understand this question, consider the question our previous theorists have been attempting to answer: "What makes X identical to Y?" This is what's known as the *reidentification* question: it asks about the conditions under which some X at one point in time is properly reidentified as Y at some other time. The answer, then, must be given in terms of numerical identity, which is about the relation something has only to itself. By contrast, the characterization

question is about the relation one has to various experiences, actions, and psychological characteristics. In other words, the characterization question asks about what makes some psychological characteristic, say—a desire, care, commitment, belief, project, goal, and so forth—*mine*, a feature of the real me. So instead of asking about the conditions under which an individual at some other time is one and the same individual as me, it asks about the conditions under which various psychological characteristics, experiences, and actions are *properly attributable* to the real me.

According to Schechtman, the characterization question is more appropriate to finding a relation between identity and ethics than is the reidentification question. One reason we might think this stems from recognizing the difficulties our theories of numerical identity have repeatedly run into when applied to our practical concerns. But another is the seemingly natural fit between the characterization question and those practical concerns. In seeking to account for anticipation, we seem to be wondering, "What makes those expected future experiences *mine*?" In seeking to account for self-concern, we seem to be wondering, "What makes those future states I'm specially concerned about *mine*?" And similarly with questions of responsibility and compensation: "What makes those actions for which I'm responsible—or those burdens for which I'm to be compensated—*mine*?" Consequently, given that these aren't questions demanding any sort of reidentification, and given that they seem to be more naturally and closely connected to our person-related practical concerns, we may well have been asking the wrong question all along. What we should have been asking, it seems, was the characterization question.

What is it, then, that makes some actions, experiences, or psychological characteristics mine? Answering this question does not require an appeal to a criterion of numerical identity. Instead, according to advocates of this approach, what it requires is an appeal to the following:

> **The Narrative Identity Criterion**: *what makes an action, experience, or psychological characteristic properly attributable to some person (and thus a proper part of his/her identity) is its correct incorporation into the self-told story of his/her life.*

This answer to the characterization question points to a process by which we *constitute* ourselves, and it involves telling ourselves a story about our lives, about where we've been and where we're going. It is via this narrative process that our identity is developed, maintained, shaped, and changed, and it involves several aspects worth discussing.

1. *Narrative identity is about what unifies a set of experiences into the life of a single person*. Instead of being about reidentification or the numerical identity relation, narrative identity is explicitly about the way in which the life of a subject of experiences becomes unified as the life of a genuine *person*, that is, it's about how the experience some five-year-old has of getting her first haircut becomes woven together with the experience an 80-year-old has of putting on a wig for the first time, such that both experiences *become* experiences that are part of the same person's life. Insofar as narrative identity is about persons, then, it privileges psychology over biology, that is, it renders the question of our essential nature—and thus the Biological Criterion—practically moot; this is explicitly a view about a certain sort of identity for a certain sort of creature, namely, *persons*. But insofar as it is also not about numerical identity, narrative identity is unconcerned with discovering the relation that makes a person at one time identical to a person at a later time at all, so it renders the Psychological Criterion practically moot as well.

2. *What renders certain experiences as unified into the life of a person is precisely the narrative that person constructs about those experiences that shapes them into that of one life*. This is a bit tricky, but the general idea is this. Experiences are not experiences *of a person* until and unless they have been incorporated into that person's life via some narrative structure, that is, until and unless they have been *appropriated* by the person as his or her own. This is, in part, because such experiences are simply meaningless unless viewed both in relation to other experiences and to the person having them. To take a simple example, suppose I have dinner with my wife. This experience is meaningless considered as some isolated event: try thinking of it as simply an image, a snapshot, of a table, food, and a woman on one side of the table. But that woman *means something* to me, as may the restaurant and the food, and what I do in thinking

back to that night is connect those events and people as one night in my life story: I remember driving to the restaurant, having a fight with my wife on the way, making up halfway through dinner (thus the explanation for that sideways smile on her face), enjoying the jambalaya so much I bought a Cajun cookbook later, and so forth and so on. The various events that made up that evening become intertwined with one another, and then with other strands of my life, via my act of narration. And something similar goes for future events: insofar as I anticipate experiencing some event, I weave it into the story of my life. If there's some party I'm looking forward to on Friday night, I "see myself" there—I may even rehearse in my head certain things I want to say to some people—and insofar as I do so I incorporate those future experiences into my life, that is, I claim them *as my own*.

3. *What constrains the incorporation of various experiences is whether they "fit together" into one's narrative, and whether they approximate reality.* First, the narrative of one's life has to be coherent; it has to make sense as a narrative. If one can't articulate some experiences or events as coherently part of one's life story, then they aren't any meaningful part of that story. Suppose, for instance, someone were to give you the following account of some past event: "I was a loving, compassionate husband, so I would hit my wife on a daily basis." This story simply makes no sense. The motives of such an individual would be unintelligible if what he says actually took place. Consequently, some aspect of the narrative must be revised: either the events in question never happened, or the person's description of who he was is just wrong.

Relatedly, one can't just make up any old story one would like to connect the various experiences that have occurred; instead, there must be some significant correspondence between the narrative one constructs and reality. It may somehow make sense for me, for example, to weave Napoleon's experiences into my own life. It may help to explain my current monomania, say, if I were to include as part of my life story that I had been the general of many victorious battles against various countries in Europe. But this simply won't be accurate or telling as part of a genuine life narrative—indeed, it is the kind of thing the mentally ill do.

Of course, these constraints will be met by degrees in actual practice: some narratives will simply be more coherent than others, some narratives will be more fractured than others. But (it's been argued) we should think of the ideal of perfect intelligibility as the aspect of our narratives to strive for, despite its probably being unattainable. What we want is that our life stories make as much sense as possible, and to that end we'll take what we can get. And the more the various elements of our lives fit together, the more defined we are as characters, and the more stable, sharp, and coherent are our narrative identities.

4. Narrative identity presupposes numerical identity. This is an important point. To adopt an account of narrative identity is not to suggest that there is no such thing as numerical identity, or that there is no point to investigating its nature, or that we as persons aren't also individuals with persisting numerical identities. Instead, narrative identity assumes the presence of numerical identity, and what its advocates maintain is just that narrative identity *accounts for our practical concerns* in a way numerical identity cannot. The real relation between identity and ethics, they claim, is that between *narrative* identity and ethics.

There are two points here. First, just as in fiction, a person's narrative is senseless unless it is the narrative of *one and the same* individual. So narrative identity is about what unifies the various actions and experiences of one and the same subject of experience into the life of a genuine person. But narrative identity is actually neutral between competing accounts of the numerical identity of that subject of experiences. In other words, the story we have told about narrative identity is perfectly compatible with the truth of either the Psychological Criterion or the Biological Criterion: what makes certain psychological elements mine, part of my ongoing biography as a person, may obtain regardless of whether or not what makes me the same individual across time is biological or psychological continuity.

The second point is that it is narrative identity, and not numerical identity, that purportedly does the real work in accounting for our practical concerns. Thus, while numerical identity is *necessary* for rational anticipation—I cannot rationally anticipate some future experiences unless I expect them to be the experiences of the individual who will be *me*—it is

not sufficient. In other words, it is not enough that some future individual will be me for it to make sense for me to anticipate his experiences, for the simple reason that he may be in a permanent vegetative state. Instead, it makes sense for me to anticipate some future experiences only if those experiences will be *mine*, that is—according to this position—only if they will be the experiences of a *person* and they fit coherently and accurately into my own ongoing, self-told life story.

It is worth saying more here about the purported advantages of the narrative view over numerical identity views with respect to our practical concerns. And insofar as the Psychological Criterion looks to have a more plausible connection to our practical concerns than the Biological Criterion, we will focus on it. So according to the Psychological Criterion, what grounds my rational anticipation of some future experience is just that that future experiencer will be uniquely psychologically continuous with me. But one might well think that, just because there's some overlapping chain of direct psychological connections between me and some eighty-year-old person, that isn't sufficient to ground my rational anticipation of his experiences. Suppose, after all, that I live entirely in the moment, flitting about to do whatever is in accordance with my strongest desire at any particular time, and that I have no ongoing projects, plans, or goals and lack self-reflection altogether. It's very hard to think of me as any kind of genuine person, or agent, at all. Instead, I am what Harry Frankfurt has famously called a *wanton*, someone who doesn't care about how his life is going or what he is to make of it.[1] Now there will be between me and my future eighty-year-old self unique psychological continuity; he will indeed be me. But this fact of numerical identity doesn't seem to provide any sort of grounds for rational anticipation on my part; after all, *what is that eighty-year-old man to me*? In no real sense will his life be mine; indeed, it's hard to think of my having any sort of life *at all*. So what rational sense can be made for my looking forward to his experiences? The narrative identity view has a real advantage over the Psychological Criterion

1 See Harry Frankfurt, "Freedom of the Will and the Concept of a Person," in Harry Frankfurt, *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1988). The adjective *wanton* means capricious, frivolous, unrestrained, arbitrary.

here, therefore, given that it can explain why unique psychological continuity isn't sufficient to ground rational anticipation; instead, rational anticipation requires the kind of personhood and psychological unity that only narrative identity delivers.

A similar account may be given for self-concern. What an application of the Psychological Criterion seems to warrant is a special sort of concern that I, a person at one point in time, may have for the person who will be me at some future point in time, such that what grounds this concern is that person's unique psychological continuity with me. So the Psychological Criterion localizes the target of self-concern to some future moment from the perspective of the localized present moment: I-now care about the well-being of I-later. On the narrative identity view, however, self-concern is a concern I, a narrative self, have for that very same narrative self—for *me*, narratively construed—and this isn't a localized kind of concern at either end; rather, it is *global*. To have self-concern is thus to care about the whole self whose life I am creating, and the Psychological Criterion cannot seem to capture this important aspect of it.

In addition, it is my self-concern that, in a way, *makes* that future mine. As Schechtman puts it, my concern for the future

> is an ongoing, active orientation that creates a kind of experience that is not present without it. The subject worrying about his future is a narrative self and not some particular moment of this self, so the effects of self-concern do not consist only in the fact that at one moment (or even at each moment) a particular *anticipated* future changes a person's present. Instead, the formation of a narrative brings into being a temporally extended subject who has this concern for her whole self. By the time someone is in the position to worry about the future he is already more than a momentary creature.[1]

Finally, even if an advocate of the Psychological Criterion were to try to adopt a more global vision of self-concern, it's unclear why one should

1 Marya Schechtman, *The Constitution of Selves* (Ithaca, NY: Cornell University Press, 1996), pp. 156–57; emphasis in original.

care about a self unified by unique psychological continuity; after all, many of those individual experiential moments may just be *irrelevant* to me and the way in which I conceive my life, so there would likely be a serious disconnect between the self I actually care about and the self the Psychological Criterion would provide me rational warrant to care about.

What all of this seems to suggest is that the kind of identity that matters for our practical concerns is narrative, not numerical, identity. If this is true, then we could move beyond the standoff reached in the last chapter by admitting that we were focusing on the wrong type of identity all along. Nevertheless, *is* this true?

Evaluation of the Narrative Identity Alternative

So what are we to make of this view? The most important advantage it has going for it is clearly practical: it provides what seems to be the best way thus far to account for the rich phenomena of anticipation and self-concern, and it does so while remaining neutral between any particular criterion of numerical identity, and so it avoids the metaphysical standoff we ran into in the last chapter. Remember, what Weirob wanted from the get-go was a criterion of personal identity that helped us make sense of these practical features of our lives, and narrative identity seems to do this very well.

Nevertheless, while it initially seems to have this significant practical advantage, there are still some real concerns we might have about the view as a whole, including:

1. *The Endpoints Problem*. Narrative identity is presented as being about the unification of various experiences, actions, and psychological characteristics into the life of a single individual, a unity that comes via the biography we construct for ourselves, constrained only by considerations of coherence and approximation to reality. It answers the question "Who am I?" by stating, as DeGrazia puts it, "*You are the individual who is realistically described in your self-narrative or inner*

story." But this construal actually allows that various non-experiential or non-psychological events, even pre- and post-personhood, could be included in one's narrative. For example, I may coherently and correctly say, "I was born prematurely," or "If I'm ever in a permanent vegetative state, you may turn off the machines keeping me alive." It seems, in other words, that narrative identity isn't necessarily about the identity of *persons* at all.

Now in itself this expansion of the enterprise does not constitute an insurmountable problem for the narrative identity view. Indeed, there are those, like DeGrazia himself, who seize on it as a way to show that narrative identity is quite compatible with, and actually presupposes, a Biological Criterion of numerical identity. A Psychological Criterion, by contrast, could not be presupposed by a narrative identity incorporating these pre- and post-psychological events.

But this way of putting it just reinforces our earlier point that, while the endpoints of one's narrative identity have to be constrained by the endpoints of one's *numerical* identity (according to narrative identity theorists), as it turns out this just isn't how self-narratives often work. A variety of events may be incorporated into my self-narrative, some of which will be contained within the arc of my biological life, but some of which won't. For example, while I will certainly want to include details of my being born into my narrative, I may also want to include details of what happened to me at various stages of fetal development. Now most Biological Criterion theorists have no problem with this, for they think that our biological lives begin around the two-week stage post-conception, at the time the possibility for twinning has passed. But why should my self-narrative start there? After all, were I to find out that during the first two weeks post-conception my embryo had indeed split and then fused back together (as may very well happen), that event would surely play a role in my self-narrative: I really could have been—and was, for a bit—a twin! And if events during that first two week period can play a key role in my narrative, why not events prior to that? Why isn't what happened to

1 David DeGrazia, *Human Identity and Bioethics* (Cambridge: Cambridge University Press, 2005), p. 83; emphasis in original.

my mother's ova, or my father's sperm, relevant? Indeed, if we are looking for *explanations* of my current identity, for what makes me who I truly am today, why aren't the events in my parents' lives (and their parents' lives) relevant to my narrative as well?

On the other end of the spectrum, while I may indeed incorporate events happening to my potential PVS-stage as part of my narrative identity, why can't I also incorporate events happening to my *corpse*-stage as part of that identity? Why can't I say, "When I die, I'd love to lie in state as did Lenin, to be viewed and adored by the masses for years on end"? Suppose, through some crazy series of events, that this in fact happens to me. Then suppose that someone comes through one day and spits on my corpse, fomenting a riot, and, ultimately, a political revolution. Surely these events are just as much a legitimate part of *my* narrative as anything else in my life. But there just is no plausible theory of numerical identity that incorporates one's years-old corpse as identical to oneself. So the endpoints of narrative identity are not in fact constrained by the endpoints of numerical identity. And this leads to the next problem.

2. *Prescriptive or Descriptive?* What precisely is the upshot of this theory about narrative identity? Is it a *descriptive* enterprise, describing the way we in fact do think of our lives, or is it a *prescriptive* enterprise, prescribing the way we in fact ought to think of our lives? As it turns out, there are problem with both interpretations. If it is a descriptive thesis, then it is false. As Galen Strawson has pointed out, there are certainly some people—"Episodics," he labels them—whose self-experience is clearly non-narrative, that is, they do not consider themselves as being selves who were there in the past or who will be there in the future. This is not to say these people are wantons either. They may be quite self-reflective, and they likely also have goals, projects, and plans. They just don't weave their various events into a single narrative arc, or claim to see meaning in some of their experiences only in relation to others. Now Strawson, who himself claims to be an Episodic, thinks such lives are perfectly normal and non-pathological, and while such a life may strike us as perhaps odd or shortsighted, it certainly cannot be ruled out as *nonexistent*.

Perhaps, then, the narrativity thesis is prescriptive, providing us with

the formula for how we *ought* to view our lives. But why would this be the case? Why should we create these inner stories, tying together the various events of our lives into a coherent narrative? One natural thought might be that doing so provides us with a valuable kind of self-knowledge and so points us to the proper targets for our self-concern, anticipation, and the like. Viewing the various moments of our lives as part of a larger biography may also provide them with greater resonance: to see my victory in a race, say, as part of the biography of someone who overcame cancer and sacrificed a great deal to be there makes the victory so much more than the feeling of pleasure one might experience in the moment; it may also serve to *redeem* a significant portion of one's life. On the other hand, though, there may very well be serious drawbacks to this sort of biographical tracing. Sometimes, to discover fully who I am is to discover some ugly truths, ones that may very well cripple me and destroy any reason for self-concern. Some of us have hearts of darkness, and it may in fact be better for us (and for those around us!) simply to leave those hearts as they are. At the very least, though, more needs to be said in favor of narrative identity, if this is the proper interpretation of the thesis. More generally, it is simply unclear *what* the proper interpretation of the thesis is supposed to be.

3. *The Practical Concerns Problem*. Narrative identity's greatest strength is in its alleged ability to account for all of our various practical (ethical and prudential) concerns. We have seen how it might do so for anticipation and self-concern. But what of our other practical concerns? We will explore how it deals with moral responsibility and compensation later on (in Chapters Seven and Eight, respectively). These are the four person-related practical concerns for which Schechtman takes us to desire an account. But as it turns out, (a) there are other concerns for which narrative identity doesn't give a good account at all, and (b) it's not entirely clear that narrative identity gives the best account of even these four.

Start with (a). There are some person-related practical concerns for which narrative identity is in fact irrelevant. The most obvious has to do with reidentification. Suppose I haven't seen you, an old friend, in ten years and so we arrange to meet at a local restaurant to catch up. I arrive

early and I'm waiting to see you come in. After ignoring several people as they walk by, I finally make a judgment about one of them and call out your name to that person. What justifies me in doing so? It can only be that I believe that person *to be you*. But this sort of reidentification is surely not a matter of narrative identity. That is, I am not making some sort of judgment about which experiences or psychological elements are truly yours or are part of the biography of your life. Instead, I'm making a judgment solely about numerical identity: I'm judging that the person I see before me now *is one and the same person as* the person I was friends with ten years ago. And something similar is true of *first-person* reidentification. When I see the photo on my mother's coffee table and say, "I was so cute back then!" I am justified in doing so solely in virtue of the fact that the photo is a picture of me, and not insofar as the experiences of that child are incorporated into my biography.

When it comes to legal—and even possibly moral—responsibility, we are likely relying on numerical identity as well. It may not matter, for instance, if some person incorporates his past criminal actions into his true life story; instead, all that may matter for the rest of us is that his DNA matches up to the criminal's, and so, for the purposes of the law, he is simply the same person as the criminal. And a similar story might be told for cases of (legal) compensation.

Now Schechtman admits as much, claiming that what a reidentification/numerical criterion can't capture are just the four practical concerns of anticipation of survival, self-concern, responsibility, and compensation. This may well be perfectly okay, but it does introduce a fracture into our formerly unified account. For we had been taking for granted that *all* of our practical concerns would bear a relation to the *same* criterion of identity, namely, the *true* one. But now it may be that there's one type of identity related to one set of concerns, and another type of identity related to a different set of concerns. As just remarked, this may not be problematic in itself, but it's worth noting now as one of the key methodological points we will consider explicitly in the final chapter.

Nevertheless, it's uncertain whether narrative identity is actually the best way of accounting for even the four practical concerns (point (b)

above). While we will consider responsibility and compensation in Part B, for now we can at least discuss the main practical concerns of this first part of the book, anticipation and self-concern. As it turns out, they may well still be, at least in part, about numerical identity. Take first self-concern. While sometimes it is indeed appropriate to say I am concerned about the fulfillment of *my* desires and goals (making the issue about the characterization question and ultimately narrative identity), at other times it seems much more appropriate to say that I am concerned about the well-being of the *person* who is me (numerical identity), where this also isn't a concern for my robust narrative self. Indeed, one key difference between narrative identity and numerical identity is that the former (typically) derives from first-person, subjective considerations, whereas the latter (typically) derives from third-person, more objective considerations. So while it's certainly possible to view my life and care about it from the inside, as the narrative self living it, it's also possible to view my life from the outside, to judge its overall value and have concern for it purely with respect to its various moments of enjoyment, say (this is how a utilitarian might view the matter; see Chapter Eight). When I do so, I'm assessing the value of these various moments *in total*, perhaps independently of how they are weaved into my narrative arc, and so to do so I must have an account of *numerical* identity that makes the person to whom they belong one and the same across time.

As for anticipation, suppose that I am terribly ill, and I am wondering if I will survive the night. Now I may indeed be wondering if it makes sense for me to expect any future experiences in the morning to be mine—this would be a kind of anticipation that asks the characterization question and whose answer depends on narrative identity. But I may also simply be wondering if there will be someone waking up in the morning *who will be me*, and this would be a kind of anticipation depending on the sense of numerical identity.

Thus, while narrative identity may be relevant to *some* of our practical concerns *some* of the time, it may not provide the exclusive account of all of the practical concerns its advocates have alleged of it. Furthermore, the theory of narrative identity itself isn't nearly as clear as it needs to be to

play the significant role it is supposed to play when applied to the world of ethics. For instance, what are the right endpoints of the narrative and what makes them so? And are we to take the theory as descriptive or prescriptive? These are difficult questions that go to the heart of the view. What are we to do, then? There is another, even more radical, alternative to explore here, one that is founded on a powerful objection to all of the various theories of identity (numerical and narrative) we have seen to this point.

Identity and What Matters

The objection stems from a series of thought experiments made famous by Derek Parfit.

> *Whole Brain Transplant Case*: I get into a terrible motorcycle accident. My body is a wreck and my heart will soon stop pumping, even though my brain is fine. As it turns out, my entire brain can be transplanted into the healthy cranium and body of my twin (whose own brain has just suffered a crippling aneurysm). The operation is a complete success, and the survivor wakes up fully psychologically continuous with me. What has happened to me?

This is obviously a version of the *Who is Julia?* case, and as we have recognized before, most people will want to say that I am the survivor here. Indeed, the only view we have run across that would unequivocally deny this conclusion is Weirob's Body Criterion. But this seems quite implausible, and as it turns out, even the advocates of the more sophisticated Biological Criterion would agree that I have survived, just as long as my brain stem—the regulator of my biological functioning—were transplanted as well. And it is not hard to see why most people would think I am the survivor here, for the resulting person would remember (or at least seem to remember) my life, carry out my intentions, persist in my beliefs and desires, have a character exactly like mine, and bear a close physical resemblance to me. Indeed, there would be no difference whatsoever, from

the inside, between what things will be like for the post-transplant person and what things would have been like for me had I simply undergone any other sort of operation and awakened afterwards. There seems no compelling reason, then, to deny that, in this case, he is me.

> *The Single Hemisphere Transplant Case*: Suppose that I have severe epilepsy, and one hemisphere of my brain is, as a result, removed to end my epileptic seizures, an operation known as a hemispherectomy. Many people have actually undergone such an operation and become eventually able to function reasonably well (with their remaining hemisphere learning how to take over the tasks previously performed by their missing hemisphere). Surely those who underwent the surgery were themselves the survivors of it—to say otherwise would be to say that the doctors performing the surgery were killing their patients, which is absurd—so there should be no doubt that *I* would still be alive in this case. But now suppose that, as in the first case, I get into a motorcycle accident and my body is about to expire. This time, however, only one hemisphere of my brain continues to function, and so it alone is transplanted into the healthy body of my twin brother. The post-operation person will once more wake up being fully psychologically continuous with the pre-operation me. What, then, has happened to me?

Once more, it seems as if I would be the survivor. If I would be the survivor in the Whole Brain Transplant Case, and I would also survive the loss of one hemisphere of my brain, there would be nothing of any additional relevance missing in the Single Hemisphere Transplant Case that would suddenly make me no longer the survivor. But if we agree with this assessment in both of these first two cases, what happens when we combine the cases?

> *The Double Transplant (Fission) Case*: Suppose that I'm in a motorcycle accident, with the usual havoc having been wreaked on my body, but that I have two healthy brain hemispheres, each of which is essentially the duplicate of the other (that is, there are no real differences between their abilities). Now suppose my two *triplet* brothers suffer aneuryisms.

> One hemisphere of my brain is thus transplanted into one brother's body, whereas the other hemisphere is transplanted into my other brother's body. After the operation, two people—call the one with my right hemisphere Righty and the one with my left hemisphere Lefty—wake up and both of them are fully psychologically continuous with me.

This case should also sound familiar. It is a more down-to-earth version of the Divine Duplication case we discussed in Chapter One. Of course, one might think that fission is even more far-fetched. Indeed, why think we can learn anything of value at all from considering something like this, something which could never actually happen?

One reply is that, at least in terms of the most important aspect of the case, fission of a kind has already occurred. The two hemispheres of our brains are connected by a bundle of fibers known as the corpus callosum, a bundle that enables the two hemispheres to communicate with one another. Scientists have found that severing the corpus callosum in patients with severe epilepsy can significantly reduce their seizures. But they've also found something else in such patients, namely, what seem to be two separate streams of consciousness. This was revealed in specially designed psychological tests. Our right hemisphere controls the left half of our body, while our left hemisphere controls the right. Once the patients' corpus callosum had been severed, though, it was as if each hemisphere of their brains communicated independently of the other. So they would be presented with a wide screen, one half of which was blue, the other half of which was red, such that each hemisphere "saw" only one color via the halves of each eye it controlled.[1] On each half of the screen was the question, "What color do you see?" One of the patient's hands wrote "blue," and the other wrote "red." And there have been other fascinating experiments and anecdotes along these lines. One patient claimed that there were times in which, when he was hugging his wife, his left hand would push her away.

---

1 The right hemisphere is connected to the right halves of each retina, and the left hemisphere to the left halves. So if this divided screen is presented too quickly for this subject to move his eyes and expose the halves of the screen to both halves of each retina, the red stimulus goes only to one hemisphere, and the blue only to the other.

So what does this mean? There seems to be, in such patients, a division of consciousness into two streams, each of which is unaware of the other. But we can easily see how this real-life case is relevant to fission, for all we are supposing in this thought experiment is that the division of consciousness came via a permanent physical separation of the two hemispheres. Whether or not such fission is ever technically possible, then, should not be a concern, given that what might have been thought to be the deeply impossible aspect of the separation—the division of consciousness—seems already to have occurred.

Consider the case, then. The first question we have about fission is exactly the same as the question we had in the Divine Duplication case, namely, what has happened to me? There are four, and only four, options:

Option 1: I survive as both Righty and Lefty? This might seem the most appealing answer, at first. After all, I would survive the Whole Brain Transplant, and I would survive the Single Hemisphere Transplant, so why not think I'd survive, just twice over, in the Double Transplant? Unfortunately, this cannot be the case, given the simple and obvious fact that there are *two* people post-fission, and two does not equal one. In other words, we want to know what has happened to *me*, one person. If we say that I am *both* Righty and Lefty, and they are two distinct persons, then we'd be forced to say that *one* person equals *two* persons, or one equals two, which is just false.

Of course, you might simply deny that Righty and Lefty *are* two persons. Instead, you might say, I am one person with two bodies and a permanently divided stream of consciousness. But making this move would cause all sorts of other serious difficulties, especially with regard to our concept of personhood. Suppose Righty and Lefty go off and live on opposite ends of the earth, and have a variety of very different experiences. It would become very difficult to continue to think of them as one single person in that case. In addition, suppose they were to play poker against one another. Would it really be just a game of solitaire? What if one shot the other in a rage? Would it be murder or suicide? And suppose, through some very strange and incestuous turn of events, they make love to one another. Would it instead simply be a case of masturbation? Once we

think about it, the negative implications of calling Righty and Lefty both me, a single person, are too overwhelming to bear.

Option 2: I survive as Righty? Here you might agree that I can at most be only one of the survivors (who are each individual persons), and then insist that I go where my right hemisphere goes. But why think this? Indeed, both hemispheres are essentially duplicates of the other, and both Righty and Lefty would be fully psychologically continuous with me, so what non-arbitrary reason is there to think that I would be Righty and not Lefty? Of course, if you adhere to the Biological Criterion, you might think that I go wherever my brain stem goes (which can't be divided), and if it goes with Righty, he would be me, but if goes with Lefty, then he would be me. But it would be quite odd to think that my entire identity would be preserved in that small bit of regulatory biology, especially when both Righty and Lefty would have their own brain stems doing precisely the same regulatory biological work as my original brain stem. Indeed, it would be almost as arbitrary to insist that my original brain stem must remain intact for *me* to remain intact as it would be to insist that either Righty or Lefty is me.

Option 3: I survive as Lefty? The same reasoning applies here: what non-arbitrary reason is there to think that I would survive as Lefty, and not Righty, given that they'd each be exactly similar to me (psychologically, at least)?

Option 4: I do not survive? As it turns out, this is our only other option, and it must be correct. I can't survive as both, and there's no reason to think I've survived as one and not the other, so I must not survive fission. (Of course, if my original brain stem went into neither Righty nor Lefty, the advocate of the Biological Criterion discussed above would agree that I do not survive as well, but for a different reason.) This is rather extraordinary, though. If I survived the Whole Brain and Single Hemisphere Transplant cases, why would a *double* success count as some sort of failure? The reason is simple: the numerical identity-relation is a one-one relation—it holds only between one thing and that same one thing—but the relation that holds between me and the post-fission people must be one-many, holding between one thing and more than one thing (if we

accept that the survivors are two distinct individuals). So the relation that holds between me and the survivors cannot be the identity relation. *I* do not survive fission.

But now we need to ask the crucial follow-up question: *does this matter?* In other words, is the fact that the identity relation is missing between me and the fission survivors an important fact? And it is here where Parfit has famously said no: identity is in fact *not* what matters in this case. To see why, consider things from the internal perspective of each survivor. Start with Righty. He'll seem to remember my life, right up to the moment in which he went under anesthesia. He'll also have my intention to go out and party tonight, he'll believe that the surgery was the right thing to do—as did I—and he'll have precisely my level of love for poker, polka, and okra. But now consider things from Lefty's perspective. He'll be exactly psychologically similar to Righty, so he too will share my memories, intentions, beliefs, loves, and so forth, in exactly the same way Righty does.

For both fission-products, then, it will be precisely as if *I* had awakened from the surgery. So if we look at things from my pre-fission perspective, everything that matters to me about ordinary survival will be preserved in both of my fission products. The only difference between this and ordinary survival will be that, whereas I would bear the relevant intrinsic relation to only one person in the ordinary day-to-day case, here I bear that relation to *two* people. Now "survival" entails identity: for me to survive some surgery, the post-surgery person has to be identical to me. But since there is no identity between the post-fission persons and me—solely because identity can obtain only one-one—I don't survive fission. But because everything that *matters* to me about ordinary survival obtains—twice over!—then what occurs in fission is *just as good as* ordinary survival.

And what *is* the relation that obtains between me and the fission-products that preserves what ordinarily matters in survival? Now it's true that there is a bit of physical continuity between us: they each have a portion of my original brain (not quite half). But of course this is important only insofar as that brain portion supports *psychological continuity* between us. What matters in ordinary survival—what I look forward to in day-to-day survival—is that the person who wakes up in my bed, say, will remember

my life, act on my intentions, see and approach the world as I would have, love and take care of the things I love and take care of, and so forth. And whether or not there is one person or there are two people who will do this is—at least to some extent—unimportant.[1] Identity, then, is not what matters; rather, what matters is psychological continuity. Call this view, therefore, the **Identity Doesn't Matter** (IDM) view.

If we accept a view like this, there will be a number of important implications for our practical concerns. We are currently focused on anticipation and self-concern, and this view does quite well in accounting for them. What we have realized is that some sort of psychological continuity relation does the best job of grounding these patterns of concern. The problem we kept running into, though, stemmed from the pairing of *identity* with psychological continuity. In the Divine Duplication case, for instance (our precursor to fission), we saw how the only way to avoid the violation of the transitivity of identity was to make up a seemingly arbitrary restriction: the relation between X and Y has to obtain *uniquely*. But this meant that, if God created only one version of me in heaven, I'd have reason on earth to anticipate survival, whereas if God created two copies of me, I'd have no reason *at all* to anticipate it, given that *I* couldn't survive.

We can see now, though, just how silly this attitude is, given that we have a very legitimate alternative: simply focus on psychological continuity *directly* and in so doing divorce our practical concerns from the identity relation itself when identity diverges from psychological continuity. It is thus not "that he will be me" that is my reason to anticipate someone's experiences or have a special concern for him; rather, it is "that he will be my psychological successor."

There are other, more radical, implications of the view, however. The most important stems from the fact that psychological continuity

1 There may indeed be some practical worries to think about were fission to take place. For instance, which one gets access to my bank account? Which one goes home to my wife? These are not minor problems! A more careful way to pitch the fission scenario, then, is just this: suppose the prospects of my fission-products would be just as good as my own (without fission). In *this* scenario, then, it should be clear that the loss of identity between me and the fission-products is not an important loss.

is made up of overlapping chains of psychological connectedness, and connectedness, unlike identity, comes in *degrees*. In other words, the relations that together constitute psychological connections—memory, intentions, beliefs, desires, cares, and character—obtain in stronger and weaker forms, relations that sometimes alter in strength from day to day. So my memories of yesterday are far stronger (and greater in number) than my memories of twenty years ago; my character now more closely resembles my character yesterday than the one I had as a child; most of my current beliefs, desires, and cares were held by my yesterday's self but not my childhood self, and so on. But now, given that many of these connections that themselves constitute psychological continuity are matters of degree, if our practical concerns are grounded in psychological continuity then it looks as if our practical concerns themselves ought to be matters of degree as well.

This could mean, for instance, that I might be rationally justified in caring less about my distant future selves, solely insofar as I expect them not to be very close psychological continuers of mine. That retirement-age self, I might think, will not care about the things I now care about, nor will he much remember my current experiences or carry out my current intentions. Why, then, should I care as much about, and sacrifice as much for, him as I do my tomorrow's self, who will be much more closely related to me psychologically? This approach might also go for the rationality of anticipation: I have more reason to anticipate the experiences of those selves I expect to be more closely psychologically related to me. And there will be, as we shall see, some very interesting implications of the view for more explicitly ethical concerns, such as moral responsibility, compensation, advanced directives, and ethical theory generally.

But the most radical implication, directly relevant to what started off our investigation in Chapter One, is that this view could allow for the rationality of anticipating the *afterlife*. Here's how: suppose God exists and has the will and ability to create, upon your death, a duplicate of you in Heaven. Now this is a very big supposition, but it at least seems logically possible. It is quite implausible, however, to say that this person

in heaven will be *you*. For one thing, he or she will not be biologically continuous with you, so the Biological Criterion rules out your identity with this person. Furthermore, the Psychological Criterion, as we have already seen, has a very hard time accounting for this case, just given the possibility of Divine Duplication or the possibility that an impatient God creates your duplicate even before you die. And given that narrative identity presupposes numerical identity, and neither criterion of numerical identity can account for your surviving your death, it looks like none of our criteria of identity allow for the possibility of such survival, in which case, if rational anticipation attaches to identity, it can never be rational to anticipate surviving one's death.

But if rational anticipation attaches to *psychological continuity*, then it could be rational to anticipate the experiences of that heavenly duplicate after all. He or she will be just like you psychologically: he or she will seem to remember living your life, persist in your beliefs/goals/desires, have a character just like yours, and so on. So what would happen, were God to make a copy of you in heaven, would be *just as good as ordinary survival*. Would the survivor be you? Probably not, although that would, on the IDM view, be irrelevant. Rather, what matters is that he or she would be psychologically continuous with you, and in light of that possible relation, it could be perfectly rational to anticipate his or her experiences in Heaven. Why wouldn't this be good enough, then? Indeed, it is this sort of possibility to which Dave Cohen refers in his final mysterious remarks to Weirob.

Evaluation of the IDM View

Of course, we know by now that no theory regarding persons and personal identity is problem-free, and the IDM view is no exception. Perhaps the most significant objection launched against it comes from Mark Johnston, who says that, while the fission case may give us a reason to divorce our practical concerns from personal identity *in the fission case*, it doesn't at all give us a reason to divorce them in all our other ordinary cases. And

let's face it: fission just never happens![1] So yes, if it were to happen, we might want to *extend* our ethical and prudential practices to deal with it, and we might ground our practices at that point on something like psychological continuity, but until that day occurs (which is quite unlikely), our practical concerns remain grounded on identity. Indeed, something like self-concern is just that: *self*-concern. It is a special sort of concern for the person who is *myself*, not the person who will be psychologically continuous with me, and that self-concern is simply part of a coherent set of self-related concerns I have simply in virtue of being a normal human being. I care about *my* friends, *my* family, and yes, *my* self, and there's no reason to think that some thought experiment about a technologically improbable procedure should have any force in undermining that very natural pattern of concern.

This is an important point, not just for the view under consideration but also for our overall project. Even if we allow such crazy cases like fission into consideration, what is the precise lesson we should draw from them? Should we really radically revise our current practical concerns in light of them? Why not instead simply preserve our ordinary concerns as the default until we actually encounter such a bizarre scenario in real life? Indeed, should metaphysical considerations more generally play *any* revisionary role in our practical concerns? These are some of the difficult questions we will put off until the final chapter. For now, however, it may suffice to reply that the fission case specifically may not be meant to cause us to revise our practical concerns at all; instead, it might be meant simply to *reveal* to us what we're already committed to given our practical concerns as they stand, namely, that these concerns in fact track psychological continuity in our ordinary lives. In other words, what the fission case may reveal to us, in dramatic fashion, is not that we *ought* to extend our patterns of concern to our psychological continuers in just this peculiar sort of case, but that in thinking carefully about the case we may in fact

1 Well, maybe something like fission happens in the very rare cases of surgical detachment of the hemispheres; but it doesn't happen in the radical form imagined to produce Lefty and Righty: the brain transplantation necessary for it just isn't technologically possible.

find that we *do* (or would) extend these concerns to both psychological continuers, that they would be successors *already* caught in the net of our ordinary natural concern. Nevertheless, more would need to be said to defend the IDM view in this way from Johnston's powerful objection.

Conclusion

We have certainly discussed quite a lot of material in this first part of the book, but what exactly is it, if anything, that we have accomplished? To see where we find ourselves, it may be helpful to retrace our route in getting here. What motivated the enterprise was a question that nearly all of us probably have: is it rational for me to anticipate surviving the death of my body? To get an answer to this question, we had to find out whether or not it was possible for *me* to survive the death of my body, and in doing so we assumed that what makes anticipation rational is personal identity, that is, in order for it to make sense for me to anticipate some future person's experiences, that future person must be me.

In our first chapter, then, we tried out the suggestions of Weirob's dialogue partners—exploring both Soul and Memory Criteria—and found, along with Weirob, that they were either irrelevant or simply unable to do the job we wanted, which was to provide a criterion of personal identity that provided a logically possible mechanism getting us from here to the afterlife. It thus seemed as if survival of death was impossible. But as it turned out, even the other "non-immortality" theories of personal identity discussed in the dialogue—the Body Criterion and the Brain-Based Memory Criterion—seemed to stumble over significant obstacles as well on the road to plausibility.

We then turned in Chapter Two to a discussion of the two most sophisticated theories of personal identity on offer, the Psychological and Biological Criteria, in order to see what their relation might be to our day-to-day prudential concerns—anticipation and self-concern—independently from the vexed question of the afterlife. What is it, we wanted to know, that makes it rational to anticipate the experiences of, and have a special sort

of concern for, that person who will be getting out of my bed tomorrow, going to my classes, fulfilling my role at work, and so forth? Once again, we assumed that the answer was, in its general form, one of personal identity: what makes it rational is that that person in the morning *will be me*. And so we set out to see which criterion of personal identity grounds this practical work. We first found that, while it did very well in accounting for our practical concerns, the Psychological Criterion ran into serious difficulties with respect to both its method (the Method of Cases) and its implications about our essence. But the theory that did fare well in *those* respects—the Biological Criterion—itself fared rather poorly in accounting for several of our key intuitions, as well as our practical concerns generally.

We turned, then, in the present chapter, to an exploration of a couple of radical possibilities. First, we considered abandoning numerical identity in favor of narrative identity, which focused on the question, "What makes me who I am?" rather than on the question, "What makes me the same person across time?" And while this move seemed to yield some fruit with respect to *some* of our practical concerns, it wasn't, at the end of the day, a very clear theory, nor did it seem distinctly relevant to other of our practical concerns.

The second radical possibility was simply to abandon the assumption that had been guiding us all along, namely, that it is personal identity that grounds rational anticipation and self-concern (and perhaps other of our practical concerns). This possibility was motivated by consideration of the famous fission case. By far the most plausible response to that case was to admit that I don't survive, but this admission, on the IDM view, wasn't supposed to bother us, given that I would still be fully psychologically continuous with both fission-products. Indeed, on this view, what matters is precisely this relation—psychological continuity—and this is the relation that does or ought to ground anticipation and self-concern, even though we had mistakenly assumed it was identity that was doing that trick. One of the IDM view's most important virtues, then, is the way in which it cuts right to the chase: we kept wanting to find a psychology-based account of anticipation and self-concern (for example, the Memory Criterion, the Psychological Criterion, and narrative identity), but we kept running into

problems constructing a theory of personal identity around the relevant psychological relations. What the IDM view does is simply deny the identity part, while preserving the psychological relations, simplifying our search profoundly. It also provides the possibility of rationally anticipating the experiences of some heavenly person, despite the fact that he won't be me, which seems to be about the most we can legitimately ask for.

There are problems with the IDM view too, of course, one that we have already seen, and others that we will explore later. But it has certainly earned a place of consideration among our other prime contenders, the Psychological Criterion, the Biological Criterion, and narrative identity. But now what? Are all four views on equal footing, or are some more plausible than others? This is certainly something for you to consider, and there is another very important question for you to mull over as you do so: what role should our practical concerns play in our exploration of personal identity? In other words, we have set aside some views as just irrelevant to these concerns, and we have noted it as a problem when some view could not account for them very well. Were we right to do so, however? Or should we instead simply try to figure out the right criterion of personal identity, say, with no regard whatsoever for how it relates to our practical concerns until *after* we have somehow independently determined what the "right" criterion is? And how would we determine that, if we make no reference to our practical concerns?

These are hard questions, and we will take them up explicitly in the final chapter of the book. For now, though, we leave these matters open as we turn to a different set of issues. Up until now we have focused exclusively on the relation between personal identity and our self-regarding reasons and concerns, discussing the issues of anticipation, self-concern, and immortality. From here on out, however, we will turn away from the issue of how to deal with ourselves to concentrate on the issue of how to deal with other people: what is the relation between personal identity and morality, we will ask, specifically *other-regarding* morality? In exploring the upcoming moral issues, we will find out a variety of interesting ways in which identity may be relevant, and in so doing we may also find some ways to answer the hard questions posed above.

WORKS CITED OR REFERENCED IN THIS CHAPTER

DeGrazia, David. *Human Identity and Bioethics*. Cambridge: Cambridge University Press, 2005.

Frankfurt, Harry. "Freedom of the Will and the Concept of a Person." In *The Importance of What We Care About*, by Harry Frankfurt. Cambridge: Cambridge University Press, 1988.

Johnston, Mark. "Human Beings." *Journal of Philosophy* 84 (1987): 59-83.

——. "Human Concerns Without Superlative Selves." In *Reading Parfit*, edited by Jonathan Dancy. Oxford: Blackwell, 1997.

Nagel, Thomas. "Brain Bisection and the Unity of Consciousness." In *Personal Identity*, edited by John Perry. Berkeley, CA: University of California Press, 1976.

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

——. "The Unimportance of Identity." In *Identity*, edited by Henry Harris. Oxford: Oxford University Press, 1995.

Schechtman, Marya. *The Constitution of Selves*. Ithaca, NY: Cornell University Press, 1996.

Strawson, Galen. "Against Narrativity." *Ratio* XVII (December 2004): 428-52.