
SELVES AND MORAL UNITS

BY

DAVID W. SHOEMAKER

Abstract: Derek Parfit claims that, at certain times and places, the metaphysical units he labels “selves” may be thought of as the morally significant units (i.e., the objects of moral concern) for such things as resource distribution, moral responsibility, commitments, etc. But his concept of the self is problematic in important respects, and it remains unclear just why and how this entity should count as a moral unit in the first place. In developing a view I call “Moderate Reductionism,” I attempt to resolve these worries, first by offering a clearer, more consistent account of what the concept of “self” should involve, and second by arguing for why selves should indeed be viewed as moral (and prudential) units. I then defend this view in detail from both “conservative” and “extreme” objections.

1. Introduction

Derek Parfit has argued that the metaphysics of persons and personal identity can have an important bearing on ethical theory. His two key metaphysical contributions involve the claims that our identity may sometimes be indeterminate, and identity is not in fact what matters for survival and our patterns of anticipation/concern; rather, what matters are certain psychological relations we now bear to certain future or past person-stages.

But what remains unclear in Parfit’s account is exactly what the moral units – the objects of moral concern – for ethical theory ought ultimately to be, given this metaphysical analysis. Are they to be persons (entities more or less corresponding to human beings), selves (entities more or less corresponding to certain stages of limited duration within the lives of human beings), or simply momentary states of experience? Parfit actually

offers each of these as a possible moral unit at various points.¹ It is the aim of this paper, however, to suggest that, if Parfit's two key arguments about the indeterminacy of identity and what matters in our identity are correct, we should take selves to be the significant moral units in any metaphysically-grounded ethical theory. Furthermore, because Parfit's own explanation of what the concept of the self involves is problematic in important respects, I hope to point out a few ways in which this concept might be made clearer and more coherent. Finally, I will defend this intermediate view from objections stemming from each of the other two alternatives. I begin with a brief exposition of the Parfitian model.

2. *Parfit's Reductionist View*

Parfit is a Reductionist about personal identity. Briefly stated, this means that he holds that the facts of persons and personal identity over time simply consist "in the holding of certain more particular facts" (p. 210) about brains, bodies and interrelated physical and mental events. If one believes the facts of persons and personal identity involve some *further* fact(s) (about, say, Cartesian Egos or souls), one is a Non-Reductionist. And Parfit strongly favors the Psychological Criterion of personal identity: X and Y are the same person if and only if between them there hold certain psychological relations and these relations have not taken a branching form (i.e., there is not a Z that bears those same relations to X that Y does).²

The two relations that together make up the identity-preserving psychological relation between X and Y are *psychological connectedness* and *psychological continuity* (together known as Relation R). Psychological connectedness holds between X and Y just in case there are between them *direct* psychological relations, such as direct memory connections, intentional connections, character connections, and connections of beliefs, desires, and goals. Further, such connections can hold to any degree. *Strong* connectedness, for instance, holds between X and Y when "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). Psychological continuity, on the other hand, holds between X and Y when there are overlapping chains of strong psychological connectedness between them. Thus, X and Y are the same person if and only if X is psychologically continuous with Y and uniqueness holds between them.

I will not here rehearse Parfit's arguments against Non-Reductionism, although I find them to be quite compelling.³ Rather, I will focus briefly on the arguments for his two most original and controversial theses, namely, that (a) some questions of identity may have indeterminate answers, and (b) identity is simply not what matters with regard to

survival and our anticipation/concern for the future. I will then briefly discuss Parfit's concept of successive selves.

a. Why Our Identity May Be Indeterminate

Consider first the case Parfit calls the "Combined Spectrum" (pp. 236–43). This spectrum involves a range of possible identity-scenarios. Suppose a scientist has a series of buttons connected up to my brain and body, and that, at one end of the possible range of cases in the spectrum, by pushing a button or two, he can destroy a few of my body and brain cells and replace them, out of new organic matter, with a few (formally) corresponding body and brain cells that are exactly similar to those of Greta Garbo at the age of thirty. At the other end of the spectrum, in which the scientist pushes all of the buttons, all of my body and brain cells are destroyed and replaced, out of new organic matter, with *all* of the body and brain cells of Greta Garbo. Now at the far end of the spectrum there is neither physical nor psychological continuity *of any kind* between me and the resulting Greta Garbo Replica. At the near end, however, there is still a great deal of continuity, both physical and psychological, between me and the person who wakes up after the operation. It would be quite similar to the everyday case in which I go to bed at night and wake up in the morning. A few body cells will have been replaced and I may have one new psychological feature. What conclusions, though, can we draw here about my survival in the *middle* range of the spectrum?

Parfit argues that the most plausible response to this case is that of the Reductionist, who claims that there are certain questions about identity that simply have no determinate answer. There is no sharp borderline in this range of cases where I suddenly cease to exist and the Garbo Replica suddenly begins to exist. True, we know that at the near end the person who exists is me, and at the far end the person who exists is a Garbo Replica. But once we have described the facts of the cases in the middle range of the spectrum, in which the resulting person has some of my cells and characteristics and some of the Garbo Replica's cells and characteristics, we have described the case as fully as we can. To ask then for a determinate answer regarding the identity of such a person is to ask a question that simply has no answer; it is an empty question. In other words, our identity can occasionally be indeterminate.

b. Why Identity Is Not What Matters

With the example he calls "My Division," Parfit illustrates why it is that identity is not what matters. In this case,

My body is fatally injured, as are the brains of my two [identical triplet] brothers. My brain is divided, and each half is successfully transplanted into the body of one of my brothers. Each of the resulting people believes that he is me, seems to remember living my life, has my character, and is in every other way psychologically continuous with me. And he has a body that is very like mine (pp. 254–5).

If one continues to hold that questions of identity always have determinate answers and that it is the relation of *identity* that matters most with regard to our special concern for ourselves, then there are four possible answers here to the question of what has happened to me: (1) I do not survive; (2) I survive as Brother A; (3) I survive as Brother B; and (4) I survive as both Brother A and Brother B. But Parfit rather persuasively argues that all four possible reactions here are quite hard to believe (pp. 256–8),⁴ again leading us directly to the Reductionist View, which claims that once we have described the case as one in which both brothers receive one half of my brain and both brothers are psychologically continuous with me, *we know everything there is to know*. On this view, the problem of my survival simply disappears. The four possibilities above are no longer seen as incompatible possibilities, only one of which could be true. Instead, they are seen to be four possible ways of describing the same situation. We may *choose* what we see to be the best description here, but that in no way means that the description we choose is the *true* description. Consequently, the important question becomes not “Who is me?” but rather “What ought to *matter* to me?”

Remember, on the Psychological Criterion of personal identity, Relation R, when coupled with uniqueness, makes up personal identity. But uniqueness does not contribute a great deal to the value of Relation R, in that uniqueness itself “makes no difference to the *intrinsic* nature of my relation to [some future me]” (p. 263; my emphasis). Thus, Relation R must be what really matters in personal identity, not uniqueness, and personal identity only matters because of the presence of Relation R.⁵

To see this point, consider what the relation is between me and the two resulting people. As stated, the two brothers are psychologically connected and continuous with me. Does this relation fail to contain some vital element? No. I *would* survive if I stood in such a relation to only one person, so how could the extrinsic factor of duplication change that relationship? The only reason *I* do not survive – the only reason I am not identical with both brothers – in the fission case is that there must be a uniqueness clause in the Psychological Criterion of personal identity to avoid the violation of the transitivity of identity. Identity is a one-one relationship and cannot hold as a one-many relationship. But in this case what has occurred to me is *just as good as ordinary survival*, in that the relation between me and my brothers is just as good as the relation would have been between me and just one brother (in a case in which my entire brain

goes into one brother, say). So even though I am not *identical* to my brothers (because of the lack of uniqueness), this should not matter to me, for I in fact still maintain with them the most important aspect of identity: Relation R. Thus, it is not identity, but rather Relation R, that matters.

c. *Successive Selves*

The final relevant aspect of the Parfitian metaphysics involves the terminology of "selves," entities which become more significant in his later ethical discussion. Selves are those person-stages united by *strong psychological connectedness*. Thus, I may legitimately refer to my ten-year-old self as a *past* self, in that there are between us very few direct psychological connections (although I am nonetheless psychologically continuous with that past self). Similarly, my eighty-year-old self will most likely be a *future* self, given that (I assume) there will be very few direct psychological connections between me now and that person I will be psychologically continuous with fifty years hence. And the parts of my life with which I am currently strongly psychologically connected are united as my *present* self.⁶ Thus, when significant reductions in the degrees of psychological connectedness occur over time, one might say, "*I* did not perform that action 20 years ago. Rather, it was my past self." And one indication that the entity in question is a past self is that the present self has an attitude of *indifference* toward that entity.⁷ In this way, different selves occasionally resemble different persons, and Parfit indicates that, at certain times and places, selves might be thought of as the appropriate objects of moral concern.⁸ Such a view would of course have significant ramifications for prudential rationality, as well as morality. For example, if my smoking now will harm my future self, and future selves are somewhat like different persons, then perhaps my smoking is *immoral* (whereas it is normally thought merely to be imprudent or irrational). It is important to note, however, that Parfit insists that talk of successive selves "is suited only for cases where there is some sharp discontinuity, marking the boundary between two selves" (p. 306). In cases where there is a gradual reduction in the degrees of psychological connectedness, a reduction *without* such sharp discontinuity (most of the everyday cases, say), talking in terms of successive selves is not warranted. "In such cases we must talk directly about the degrees of connectedness" (p. 306).

3. *Selves and the Morally Significant Units*

I believe the concept of the Parfitian self as it stands is too vague to play a decisive role in a discussion of morality. For one thing, it remains

unclear just when and why we are warranted in talking in terms of successive selves. For another thing, it remains unclear just what role is played by psychological connectedness – the unity-relation of the Parfitian self – in terms of what matters, both in my survival and in my anticipation for the future. What I will propose is that when there are severely reduced degrees of psychological connectedness between two stages of a life, *regardless of the cause*, then a person can legitimately talk in terms of successive selves and refer to a past self or selves. Thus, even if changes in character, memory, beliefs, etc. occurred gradually, as long as there have been, or will be, *significant* reductions in connectedness, persons may be warranted in referring to past or future selves. In addition, I will show that in an account of what matters in survival and anticipation, psychological connectedness is far more important than continuity and, as a result, selves ought to be the objects of moral concern. Spelling out these changes will result in a more plausible, coherent, and viable form of Parfitian Reductionism, a form I will call Moderate Reductionism.

First, imagine the following “Revised Combined Spectrum.” Instead of undergoing a range of operations in which at each distinct point of the spectrum a person wakes up with a certain combination of my cells and characteristics and those of a Garbo Replica, I am given a pill that will cause these changes to take place *gradually* over the course of a day. So every few minutes a few cells and characteristics are replaced with those of the Garbo Replica. The difference between this case and Parfit’s Combined Spectrum is that, in this case, I am psychologically (and physically) continuous, but not psychologically connected, with the end-of-the-day Garbo Replica, whereas in Parfit’s case there is neither continuity nor connectedness obtaining between us. What are we to say here?

At the near end of the spectrum, the survivor is, as in Parfit’s version, obviously me. But what about at the far end of the spectrum? Here it *seems* just as obvious as it did in Parfit’s Combined Spectrum that the resulting person is not me, but is a different person, namely a Garbo Replica, *even though that person is continuous with me both physically and psychologically*. Therefore, because the person toward the end of the day is continuous with me but has no significant psychological connections to me, that person can be construed, according to Parfit’s guidelines, as my future self. But that future self seems also to be, in effect, *a different person*. Further, nowhere along my Revised Combined Spectrum does there appear a point of sharp discontinuity. Rather, the Garbo Replica is gradually continuous with me. But how can this be? After all, the Psychological Criterion of personal identity would hold that the Garbo Replica in my Revised Combined Spectrum *is* the same person as me because of the non-branching psychological continuity involved. Does this mean that the Psychological Criterion is insufficient as a criterion for personal identity?

I do not think that this case calls for a new criterion of identity. Rather, it points out the vagueness of a central feature of Parfit's picture, and it also points out the importance of connectedness over continuity with regard to ordinary survival and future-related concern. The vagueness involves the unclearly expressed relation between "selves" and "persons." Are selves merely analogous to persons or are they equivalent to persons? Are they a subset of persons, or are they to replace talk of persons? Parfit himself gives very mixed signals with regard to this issue. I here wish to offer an account of "selves" that stays true to the general Reductionist thesis but clears up the vagueness involved. Doing so will also allow us to see the reasons for the importance of connectedness over continuity in an account of what matters in survival and our anticipation/concern for the future.

First of all, there are two senses of the phrase "same person." One is the formal and strict sense, the sense in which we might say the *logic* of identity is at issue. Any criterion used in determining that X is the same person as Y in this sense must be able to handle all the requirements of the identity-relation, i.e., it must be transitive and non-branching, and it must yield an all-or-nothing answer. On the other hand, there is a looser, more popular sense of "same person" involving what might be called the *nature* of identity, as in the explanation, "I am no longer the same person as that cute and adorable little boy you once loved." In its nature, identity involves degrees.⁹

Now I suggest that the language of successive selves is to be used to capture this second, popular sense of "same person." It is a far looser sense than the formal sense in that it fails the tests of transitivity and all-or-nothingness, and thus it *could not be a criterion of identity over time*. But it does capture something important with regard to the internal life of persons and the attitude one may have with regard to one's past and future. For one may look upon certain parts of one's past and future with a kind of *indifference*, an attitude analogous to one's attitude toward other persons, and an attitude reflecting the lack of connectedness one may presently bear to those parts of one's past or future.¹⁰

So how does all of this relate to the case of my Revised Combined Spectrum? First, this example shows that one can be warranted in using the language of successive selves *regardless of the cause involved* in the reduction of the degrees of psychological connectedness. Parfit warns that we shouldn't use such language unless there is some sharp discontinuity involved. But why not? It is the *fact* of reduced degrees of connectedness that is at issue, not the *cause* of such reduced degrees. The person at the far end of my Revised Combined Spectrum is exactly similar in every way to the person at the far end of Parfit's Combined Spectrum, as is the person at the near end. What about the discontinuities involved in Parfit's case could possibly make the resulting persons any different from those in my case, where the transition is gradual and smooth?

Second, according to Parfit, Relation R plus Uniqueness equals Personal Identity (p. 263). In his Combined Spectrum, given the formal aspects of the case, the person at the far end of the spectrum is not me, simply because she is not continuous with me. But in my Revised Combined Spectrum, though there is as well no connectedness between me and the Garbo Replica, there *is* continuity, as well as uniqueness. And so we end up with the odd conclusion that, given Parfit's Psychological Criterion of personal identity, the Garbo Replica in his Combined Spectrum is *not* me, while the Garbo Replica in my Revised Combined Spectrum *is* me, and is, further, because of the reduced degrees of connectedness, my future self. But she is identical to me only in terms of the *logic* of identity.

The only thing that differentiates the two cases is the presence or absence of psychological continuity. Is this an important difference? Everything else about the two cases is exactly similar. In both cases, at every stage of the process, the two resulting people will be exactly similar. And they will both, at every stage, bear exactly similar degrees of psychological connectedness to me. Thus, the presence of psychological continuity in my Revised Combined Spectrum seems to bring nothing more to the table than personal identity in its logical sense, and this presence seems to provide little or no difference, in terms of the internal and external nature of the resulting persons, between my Garbo Replica and Parfit's Garbo Replica. And something that makes little or no difference cannot be something that makes an *important* difference.

Further, psychological continuity is what provides personal identity over time, according to the Psychological Criterion, but if personal identity is not what matters, why should continuity itself matter? Parfit maintains that *both* relations of Relation R matter, and he claims that he knows of no argument for the claim that one relation matters more than another (p. 301).¹¹ But if this is the case, and continuity turns out to have no real importance in terms of what matters, then why should we regard *connectedness* as being of any importance either? In what follows I will show, contra Parfit, that of the two relations in Relation R, connectedness is what matters in an account of survival/anticipation and, further, that continuity matters far less than such connectedness.

Parfit uses the case of "My Division" to show that what matters in survival is not personal identity, but rather Relation R. He argues for this position by considering my relation to each of the two resulting people. Even though those two people are different persons, the relation I bear to them both does not fail to contain any vital element that is contained in ordinary survival, so division is just as good as ordinary survival in terms of this intrinsic relation. Thus, what matters in survival is Relation R.

But we can now ask the same question with regard to the two relations involved in Relation R, namely, what is my relation in my Revised

Combined Spectrum to the resulting Garbo Replica, and does this relation fail to contain some vital element contained in ordinary survival? The answer is that my relation to the Garbo Replica is one of psychological (and physical) continuity, but this relation *does* fail to contain many vital elements contained in ordinary survival: the Garbo Replica would have no memories in common with me, would have very few, if any, of my own beliefs, would have a distinctly different personality, etc. In other words, she would bear an extremely low degree of connectedness to me. But ordinary survival requires a much higher degree of connectedness – it requires, in fact, what Parfit has termed *strong* connectedness – and here we have a case in which what has happened to me (the me at the beginning of the day) by the end of the day is just as bad as ordinary death because of the severely reduced degrees of such connectedness. Parfit's example of why Relation R is what matters in survival (the case of "My Division") is convincing *precisely* because of the presence of strong connectedness between the two resulting people and myself. If the resulting people were somehow continuous, but very weakly connected, with me, it is hard to see how Parfit could possibly claim that what happened was just as good as ordinary survival. So in terms of what matters in survival and anticipation/concern, strong psychological connectedness seems to be the necessary *and* sufficient element, the element that matters far more than mere continuity. There is ordinary survival, or something just as good as ordinary survival, just in case there is strong connectedness between two stages of a person (or persons).

Now the applications of the Revised Combined Spectrum case should be obvious. Instead of imagining the changes to take place over one day, imagine them taking place over the course of eighty years. Here we have something more closely akin to the normal lives of persons. Now certainly it is the case that persons typically do not change genders and psychological characteristics to the extent completely analogous with the persons in the Revised Combined Spectrum. But a great lessening of the degrees of psychological connectedness *does* normally take place (and one's physical characteristics change a great deal as well) over such a span, and my remarks would seem to apply in such normal cases as well. What matters in my everyday survival is connectedness, not continuity, and as that is what matters, that is also the relation that ought to guide my self-related concern for the future. In so far as connectedness defines the relevant boundaries of who I am, it should also define the relevant boundaries of my prudential reasoning.

Finally, we can now bridge the gap between successive selves and morality. If connectedness is the relation that matters in survival, then a lack of connectedness (or a severely reduced degree of connectedness) would be just as bad as ordinary death. But if that is the case, then a future self, because of such reduced connections, will be *like a different person* in

relation to my present self. It will be as *if* I have died and a new person exists, one who is physically and psychologically continuous with me, but who is more or less indifferent to me.¹² Now in terms of the logic of identity, in normal cases I am the same person at age eighty as the ten-year-old person with whom I am continuous. But in terms of what *matters*, in terms of the *nature* of identity, the eighty-year-old self will be different from the 10-year-old self in much the same way that I am different from some other people. And if our conception of morality involves a determination of how I ought to treat other people, then it would also seem to involve a determination of how I ought to treat entities that, for all intents and purposes, are *like* other people. Consequently, on this more consistent and plausible view – the view I call Moderate Reductionism – the scope of morality is widened to target selves as the morally significant units.¹³

4. *Objections*

To this point, the thrust of my argument has been to show that, if Parfit's basic claims about the occasional indeterminacy and ultimate unimportance of personal identity are correct, then it is psychological connectedness that matters in terms of my ordinary survival and anticipation/concern for the future. If psychological connectedness is the only relation of any significance in personal identity, and it provides the unity-relation for the self, then selves ought to be viewed as the significant *prudential* units, i.e., they ought to be thought of as the basic units of prudential reasoning. When I reason about my future and what it is in my best interests to do, I may, for example, be justified in discounting the interests of my anticipated future self, in so far as that self will not be connected to me in important respects.

But I have suggested above that selves ought also to be considered the significant *moral* units, i.e., they ought to be thought of as the basic objects and subjects of ethical theory: objects, in so far as they are the recipients of benefits, burdens, pleasures, pains, and the like; subjects, in so far as they are the moral agents, performing actions for which they are morally responsible, having and acting upon morally relevant interests, exercising autonomous choices, etc. But just because selves ought to be the basic prudential units, why exactly should we think of them as the basic *moral* units (or, perhaps better, the morally significant metaphysical units)? To this point, I have merely offered a fairly straightforward assumption: if selves are the only "person-units" that matter for prudential concerns, then it seems natural to posit them as the only units that matter for moral concerns. After all, shouldn't we target those metaphysical entities in our ethical theories that play the central role in our theories about prudential rationality?

Unfortunately, things are not so simple. For even if selves are the primary metaphysical units in prudential matters, there may be good practical reasons to target some other metaphysical units for purposes of ethical theory. In other words, even if psychological connectedness is the unity-relation that matters for ordinary survival and our patterns of anticipation/concern, some *other* unity-relation may be what matters for moral purposes.¹⁴ Let us continue, therefore, to assume that general Parfitian Reductionism is true. If so, then, as mentioned at the top, there are three possible metaphysical units that one might target as being the morally significant units: (a) persons (where psychological continuity is the relation that matters for moral purposes), (b) selves (where psychological connectedness is the relation that matters for moral purposes), or (c) momentary experiencers (where neither continuity nor connectedness matter for moral purposes). As I have already labelled the advocate of (b) a Moderate Reductionist, let the advocate of (a) be labelled a Conservative Reductionist and the advocate of (c) be labelled an Extreme Reductionist. In what follows, I will show the superiority of the Moderate view over both the Conservative and Extreme versions of Reductionism. In so doing, I will focus on the possibility of metaphysical units other than selves *qua* moral *agents*, as agents are the entities of predominant concern in contemporary ethical theorizing.

a. Extreme Reductionism

First, consider the view of the Extreme Reductionist, who holds that, while certainly personal identity does not matter morally, neither do any other diachronic psychological or physical relations. The argument for such a position might run as follows: the only unity-relation that could matter for moral purposes is a deep metaphysical Non-Reductionist unity-relation such as that provided by a soul or Cartesian Ego; there are no such entities (given the truth of Reductionism), so there is no other relation that matters. Consequently, persons are completely (metaphysically) disunified, and any ethical (or, for that matter, prudential) view that purports to justify certain rules or distributions based on the (metaphysical) existence of enduring individual unities is false. The morally significant units should then become simply the states people are in at particular times, and so an ethical theory that focused on *these* “units” would be the most plausible.

Consider, for example, a bizarre case involving a question of moral responsibility. Suppose that we have available to us a new form of “travel” called teletransportation.¹⁵ When I step into the teletransporter on Earth, the machine scans my body while destroying it and maps my entire cellular structure. It then sends (faxes?) that blueprint to Mars, say, where

an exactly similar replica of me is constructed in a matter of seconds out of a bank of cells kept there. So the person walking out of the teletransporter on Mars will be exactly similar to the person walking into the teletransporter on Earth. Occasionally, though, the teletransporter malfunctions and, even though an exactly similar person still walks out of the machine on Mars, the person on Earth is not destroyed but is instead made gravely ill, only to die a few days later. At this point, the original person on Earth may even talk to its Replica on Mars before dying. Call this the "Branch-Line Case."¹⁶

Now suppose that I have committed some immoral (and illegal) action on Earth and I intend to teletransport to Mars to make my getaway. Unfortunately, though, the teletransporter malfunctions and my body on Earth is not destroyed. Nevertheless, my Replica – call him "Backup" – walks out intact on Mars. He is then immediately apprehended and charged with my crime. He is told that because he is strongly psychologically connected to me, and because he is just like me in every way, he is morally responsible for my deed. Of course Backup protests, claiming that it is unfair. After all, *he* did not commit the crime; *I* did. And even though we are strongly psychologically connected to one another, he and I are simply not the same person; we are not identical to one another (given that uniqueness does not obtain between the pre-teletransportation person and Backup).

Parfit claims that most of us would side with Backup here, because "[w]e would believe that, in the absence of personal identity, these psychological connections cannot carry with them desert or guilt."¹⁷ Replying in this way would reveal us to be Non-Reductionists, because personal identity would remain more important than the various connections involved between Backup and myself. What we would believe to be missing in the relation between Backup and myself is the "further fact," the fact which we must believe is necessary for moral responsibility. But because there *is* no such fact – it is *always* missing – we are forced to draw another conclusion: "[n]o one ever deserves to be punished for anything they did."¹⁸ Furthermore, we must insist that the only remaining metaphysical units significant for questions of moral responsibility are momentarily existing experiencers: with the loss of the "further fact" of identity, we lose our justification for targeting agents of any more duration.

While I believe this argument ultimately fails,¹⁹ its concluding position – that the only metaphysical units we ought to target for moral agency are momentaristic experiencers – is certainly not independently *prima facie* incoherent or entirely implausible,²⁰ and for our purposes that is all that matters. For one might hold without incoherence or contradiction that, despite the existence of various psychological and physical connections that may obtain between person-stages, these relations are entirely irrelevant in a determination of moral agency. Without the further fact of

identity, perhaps the only moral agents we have are momentarily existing experiencers.²¹

But is this extreme view a truly viable alternative to Moderate Reductionism? After all, what would it actually mean to have these momentary experiencers be moral agents? There are in fact two serious practical problems with such a proposal. If we take as a given that moral agents are entities having both interests and reasons for action, we shall see that the momentary experiencers targetted by Extreme Reductionism seem to have neither.

Consider first the having of interests. Moral agents are, at the very least, entities concerned with advancing their own interests, and, if this is the case, they must have interests they are concerned to advance. But what interests could momentary experiencers be concerned to advance? Most of our ordinary interests are necessarily tied to our having pasts and futures. For example, I may have an interest in rectifying the mistakes of my past or improving my poker game in the future. But these are interests a momentary experiencer simply could not have, for it would have neither a past nor a future.

Perhaps, then, momentary experiencers simply have interests regarding their present existence. The most obvious referents for such interests would be pleasure and pain, e.g., "I-now want pleasure-now; I-now want no pain-now." But even this possibility is problematic, for as David Brink points out, the most plausible version of hedonism views pleasure and pain as functional states: "[P]leasure is a mental state or sensation such that the person having it wants it to continue and will, *ceteris paribus*, undertake actions so as to prolong it, whereas pain is a mental state or sensation such that the person having it wants it to cease and will, *ceteris paribus*, take action to make it stop."²² But if this form of hedonism is true, then momentary experiencers cannot even have these sorts of interests, for their referents – pleasure and pain – are already states of enduring entities. To experience pleasure or pain, on this view, is to be an entity with a future beyond the immediate present, and so momentary experiencers, it seems, could not even have interests regarding such states.

Furthermore, it is questionable whether or not any form of *hedonism* is even the correct theory of welfare. It seems more plausible to suggest that what is good for me extends beyond mere pleasure and pain to what kind of *person* I ought to be, and on this view of my welfare, it is obvious that I must be an enduring entity, one with a character capable of being developed over time. As Brink puts it, this view of welfare "implies that it is temporally extended beings ... who are the bearers of interests."²³

Now the advocate of Extreme Reductionism may respond in one of three ways. She may: (a) provide an alternative account of welfare that somehow makes sense of momentaristic agents having interests; (b) maintain that momentaristic experiencers are the only significant moral agents,

despite their inability to have interests; or (c) deny agency to *any* entities, including momentaristic experiencers. Responses (b) and (c) seem quite implausible, however. Response (b), for instance, maintains that no moral agents have interests. But then what exactly could agency consist in? Without interests, an agent would be without desires, motivations, or reasons to act of any kind. How then could this entity, the so-called target of ethical requirements, be an agent, a moral *actor*, without such capacities? This response wreaks havoc on our most basic conception of agency, and without serious arguments for a radical revision of that conception, it remains a non-starter. As for response (c), I suppose one could without incoherence deny that there are moral agents, but then I fail to see why one would tout momentaristic experiencers as having any *moral* significance whatsoever. After all, if there are no moral agents, then I fail to see the significance of morality in general. Without there being agents capable of choosing among possible courses of action and then acting on those choices, ethical permissions and restrictions would seem entirely irrelevant. Perhaps this respondent would want to maintain that these entities are significant for morality in so far as they can still be the passive recipients of benefits and burdens (moral *objects*, as I have previously termed them), but without there being *some* moral agents, who exactly would *distribute* such benefits and burdens? At the very least, much more would need to be said here.

Suppose, then, that our Extreme Reductionist goes with option (a) and offers us an alternative account of welfare in which these momentaristic agents have interests. Even so, it is highly doubtful that such interests would be enough to provide these agents with any *reasons* for action. After all, even if I have certain interests, what reason would I have to act at all, at any moment, if I did not think it would be *me* that would live through, and benefit from, that act?²⁴ A sense of one's *own* future is essential for having reasons to act now. As Bernard Williams puts it, what, in a very real sense, give me "a reason for living"²⁵ are my present projects, and I cannot even understand them as *my* present projects unless I also understand them "as the *projects of one [person] who will ... change.*"²⁶ Thus, for the projects which give meaning to my life to be comprehended as mine, I must conceive myself as an enduring entity, an entity propelled into the future by those present projects.²⁷

Of course, the Extreme Reductionist holding to response (c) may admit this necessary self-conception while maintaining that moral agents are still really not such enduring entities.²⁸ We might, then, permit our momentary experiencers to deceive themselves about their metaphysical existence to best preserve their psychological well-being, and our ethical theories will then, perhaps, target these non-existent, self-conceived "enduring" entities as being of moral significance.

But this response simply concedes defeat, for it no longer maintains that momentary experiencers are in fact the moral units. Instead, it claims that metaphysical units of longer duration must be the targets of ethical theory for pragmatic reasons. Indeed, we can see this point most clearly when we examine the self-aware Extreme Reductionist who advocates such a position, for why would she have any reason herself to act if she knows the “truth” about her own non-enduring existence? What reason would she have for promoting (or, in this case, concealing) Extreme Reductionism or doing anything at all? To advocate Extreme Reductionism is to concede the requisite self-conception of enduring agency, and if this is a self-conception we are to take seriously for moral purposes, then we cannot plausibly maintain that momentary experiencers are to be the metaphysical units of moral significance.

b. Conservative Reductionism

Extreme Reductionism holds that the temporal duration of selves is too long for them to be moral agents. However, as we have just seen, this view is too implausible in its own right to present a viable alternative to Moderate Reductionism. A more serious challenge, though, comes from the opposite side of the spectrum, a challenge that involves, among other things, the claim that the temporal duration of selves is too *brief* for them to be moral agents. Instead, even if general Reductionism is true, we should focus on what we normally take to be *persons* as the metaphysical units of moral significance. This is the view I have called Conservative Reductionism.

The Conservative attack on Moderate Reductionism that I will discuss is twofold. First, it contends that selves should not even be considered the entities of significance for *prudential* reasoning, i.e., psychological connectedness is not the relation that matters in terms of my anticipation/concern for the future. Rather, it is psychological continuity that is the relation of significance here. Further, selves should also not be considered *moral* units for practical reasons: taking them seriously as such would be an ultimately arbitrary choice that would involve admitting an unnecessary and confusing proliferation of agents into our ethical ontology. I begin with the first objection.

Parfit has argued that both relations of Relation R – psychological connectedness and continuity – matter in terms of survival and anticipation/concern (prudential reasoning). I have argued instead in the first part of this paper that it is actually connectedness – the unity-relation for selves – that matters in this arena; continuity is unimportant. But of course it remains possible to argue the opposite, viz., connectedness is not what matters; rather, it is *continuity* – the unity-relation for persons – that matters for prudential reasoning. David Brink does just this.

In claiming that both relations matter, Parfit asserts that connectedness matters to us because we are averse to losses of it. Thus he writes that we would regret sudden losses of memory, desires, or character traits, even if such losses would not involve a break in continuity (i.e., there would still be overlapping chains of connectedness involved). We value such facets of our present psychology, and “[w]e will want these *not* to change. Here ... we want connectedness, not mere continuity” (p. 301; emphasis in original). But Brink maintains that just because we are averse to this sort of loss, that does not require us to take Reductionism as involving the importance of both continuity *and* connectedness.

Consider a case in which I will truly be averse to the changes Parfit has in mind. This is a case Brink refers to as “corruption,” where a substantial change in psychology makes me a “less attractive person.”²⁹ A natural way to explain this attitude, claims Brink, is that I am averse to the disvalue of the psychological traits I will be getting and *not* to the loss of connectedness in itself. This explanation is buttressed by the fact that I would not be averse to a loss of connectedness that involved a change in psychology I would consider to be an improvement, “provided that I am responsible for the change in ways that establish psychological continuity.”³⁰ The general problem, then, is this: a version of Parfitian Reductionism that emphasizes the importance of connectedness (either on an equal footing with, or as more important than, continuity) cannot make sense of why we would ever have any reason “to improve ourselves in ways that involve significant psychological changes.”³¹ But self-improvement seems to be a clear-cut case of prudential rationality, and in so far as my Moderate Reductionism seems to undermine justifications for criticism of failures to improve oneself, this implication would seem to constitute grounds for skepticism that connectedness is indeed what matters. What we are left with, then, is the Conservative Reductionist view that it is *continuity* that must matter, for continuity at the very least can accommodate our most basic intuitions about the rationality of self-improvement: it remains rational for me-now to anticipate surviving as, and reaping the benefits of, my new and improved future-me, given that it will still be continuous (albeit not strongly connected) with me-now.

The case of self-improvement and connectedness, however, remains far more complex than Brink allows. First of all, more needs to be said regarding just what psychological changes would be involved in a true case of psychological disconnectedness (or a severe reduction in the degrees of connectedness). Consider, for example, a case in which I am about to undergo severe memory loss, wherein many, if not all, of my lifetime experience-memories will disappear (perhaps I am in the early stages of Alzheimer’s). I would certainly be averse to this occurrence. But is it to the loss of connectedness that I am averse, or is it, as Brink would have it, that I am averse to the new psychological make-up I will soon

possess? It seems clear here that, if I am averse to my oncoming psychological profile, I am so in so far as that new psychological profile will be the direct consequence of a loss in memory, a relation to the past that provides me with a recognizable present. Indeed, how can I know who I am if I no longer know who and where I have been? The psychological profile I will be averse to is one of confusion and disorientation, and I can easily recognize this state to be the direct result of a loss of an important strand of psychological connectedness, and in so far as the loss of connectedness is at the root of my aversion to the oncoming psychological profile, it is to the loss of connectedness that my aversion is (and ought to be) directed.

Similar remarks would also seem to apply to losses of other strands of connectedness, including significant losses of present desires, beliefs, and character traits, which contribute to my being a "less attractive person" against my will. The key in all of these instances will be just how significant the changes are. I will be averse to those losses of psychological connectedness that will result in what I foresee to be a self *with which I can no longer identify*. If the changes in these sorts of characteristics are going to be great enough, I will no longer see them as resulting in a self that bears any significant relation to me-now, and this is precisely an aversion to a loss of psychological connectedness.

Nevertheless, these are not really the cases Brink seems most interested in. Rather, he would have us focus on cases in which the changes wrought (a) result in my being a more *attractive* person, and (b) are brought about *by me* (as opposed to the Alzheimer's case, in which I am made worse off by changes over which I have no control). Can Moderate Reductionism make sense of *these* cases?

I have argued elsewhere that it can.³² The central point is this: two person-stages may remain unified as parts of the same self, *despite* there being significant psychological differences between the two. Great psychological change is compatible with the preservation of a single self. How so? Suppose at age thirty I am an uneducated alcoholic who has never been able to commit to a relationship, keep promises, tell the truth, etc., and one night I have an epiphany: I don't want to be this type of person anymore. I intend at that moment to have my life turned around by age forty. So I enroll in Alcoholics Anonymous, I attend college, I strive for honesty, I cultivate a new set of friends, and so forth. By the age of forty, I seem to have achieved my goals: I am truly an improved person. Now Brink thinks that the only Reductionist view that can adequately account for our intuitions that I have improved *myself* is the view in which psychological continuity is the relation that matters and the metaphysical units of significance for prudential reasoning are *persons*. But this is simply not the case, for even though my forty-year-old stage is rather psychologically different from my thirty-year-old stage, there are at least three

quite significant psychological connections that remain between us: at age forty (a) I retain many memories of that thirty-year-old stage; (b) I am in the process of satisfying an ongoing desire to become a better person; and, most importantly, (c) I am still tied to that thirty-year-old stage by the ongoing *intention* that has structured my life since then. This last is a relation Christine Korsgaard has labelled "authorial connectedness":³³ what in part unites the two stages is the ongoing *intent* to unite them. At age thirty, I want to enjoy the benefits of being the new and improved me, so I have set forth on projects that will bring about those benefits, and I have done so in such a way as to maintain enough psychological connectedness that I can, at age forty, reap what *I* have sown. I have, in effect, united my *self* through my project of self-improvement.

What I am suggesting, then, is that, in certain cases, ongoing desires and intentions may be enough to provide the strong psychological connectedness required for two person-stages to be unified as one self and to justify the prudential reasoning involved in self-improvement. Further, my *identification* with some past stage of myself reflects such connectedness.³⁴ I may, for example, be proud at age forty of the will and determination of that gutsy thirty-year-old – proud, that is, of *myself*. I can still identify with that past stage, still understand his motives and his confusion and his anxiety and his hope as being *mine*, and this ability reveals my ongoing unity with him.

Notice, however, that this identification is most likely *lacking* when I, at age forty, consider my relation to my *twenty-five-year-old* stage, when I was an uneducated, misanthropic drunk with no desire or intention to change. In this case, it makes sense, I think, to talk about that self as in fact a *past* self. For while I may retain some memories of that entity's experiences, they no longer seem to be *my* memories, in that they involve the experiences of a person I can no longer recognize, understand, or identify with. No significant intentions or desires bind us together, and, given the numerous other psychological changes that have occurred, we are very weakly psychologically connected. In fact, my epiphanic experience at thirty and my subsequent intentions to change that serve to unite my thirty- and forty-year-old stages are perhaps best described as marking the boundaries between my two selves. What I did at age thirty was to *sever* my connections to that past self, so that I would at age forty feel about him as I in fact do: as a stranger. And so just as there can be authorial connectedness uniting two stages of a person as one self, there can also be authorial *disconnectedness*, an act of will serving to disunify different parts of one's life.³⁵

One last example will serve to reinforce the point that it remains connectedness that matters in cases of prudential reasoning about self-improvement. Suppose that, instead of the process described above, I am offered a pill at age thirty that will, literally, produce all the changes I want

overnight. But not only will the person who wakes up have the character traits I now want, he will also have a new set of experience-memories, a new set of desires, and a new set of intentions. Not only will he have no real memories of my present life, he will also not be aware of having formed any intention to improve himself. Should I then take the pill?

There are two important points to notice about such a case. First, we can assume that the person who awakens the next morning will most definitely be an improvement over me, both from an objective standpoint and from my own subjective standpoint. He will not have my own negative characteristics (and certainly will not be bogged down by wallowing in my own sordid past), and he may have many more positive traits than I currently possess. The second thing to notice is that the person who wakes up will be psychologically continuous with me. We can imagine the changes taking place gradually overnight, where at each moment the resulting person is strongly psychologically connected to both its predecessor and its successor. So these overlapping chains of connectedness provide continuity between me and the morning person (this case should begin to sound quite familiar by now). Again, should I take the pill?

I think it rather clear in this case that I would be rational to be averse to taking such a pill, even though the resulting person would be both continuous with, and an improvement over, me (even according to my own standards of improvement). What was present in the first case (where the changes took ten years) that provided both unity and justification for embarking on my self-improvement journey is precisely what is missing in this case: psychological connectedness. I am certainly not averse, here, to the new psychological profile that will result; rather, I am averse to changes that will, in effect, leave "me" behind. Without any sort of relations that directly connect me to that future person, I am unable to identify with him, nor he with me, and so I now will be unable to reap what I sow. Without such connectedness, the case becomes less like self-improvement and more like self-sacrifice, or, perhaps better, suicide. In fact, this seems to be a case Brink's own view cannot handle, for he would have to maintain that, given the continuity that will prevail, I would be irrational to be averse to taking the pill. But in so far as it seems perfectly rational for me to be so averse, it is still psychological connectedness that matters in my prudential reasoning.

Up to this point I have merely re-established the main point of the first part of the paper: against the challenge of Conservative Reductionism, I have shown that Moderate Reductionism's emphasis on connectedness can handle cases of self-improvement in an intuitively satisfying way. But what of the final step, viz., the move from selves as the metaphysical units of prudential significance to selves as the metaphysical units of *moral* significance? Perhaps there remain practical problems with thinking of selves as moral *agents*. It is to a discussion of this final objection that I now turn.

Several authors have put forward some version of the objection that taking talk of selves literally for moral purposes is pragmatically unfeasible.³⁶ I will here focus solely on the most general of these objections, viz., Brink's claim that literal talk of selves requires an unnecessary and confusing proliferation of moral agents, leading to hopeless indeterminacy in moral discourse.³⁷ The upshot of such an objection is that we should instead target persons – for which psychological continuity is the relation that matters – as the default moral agents for pragmatic reasons.

The problem alleged to be associated with successive selves as moral units is that such selves will overlap. So temporal slices of one self will be slices of at least one other self as well. Consider, for example, the selves of Marie. Say that strong psychological connectedness holds between her twenty- and twenty-five-year-old stages (but it does not hold between her nineteen- and twenty-five-year-old stages, nor does it hold between her twenty- and twenty-six-year-old stages). Let us then call this unit of her life Self 1. But of course her twenty-five-year-old stage is strongly connected with her twenty-six-year-old stage, which is also strongly connected with her twenty-one-year-old stage, so let us call the twenty-one- to twenty-six-year-old unit Self 2. Finally, consider Self 3, which is the strongly connected unit stretching from her twenty-two-year-old stage to her twenty-seven-year-old stage. To visualize matters, consider *Figure 1*.

Now consider the twenty-three-year-old Marie (23Y). That stage is a member of three different overlapping selves, or, as the Moderate Reductionist would seem to have it, three different agents. But in so far as agents deliberate and have reasons for action, *whose* reasons are we to target here, those of Selves 1, 2, or 3? Sure, the reasons themselves can be identified, but given that such reasons are to be related to practical deliberation, “[w]hose practical deliberation is in question?”³⁹ There seems to be no non-arbitrary way for the Moderate Reductionist to identify the agent at issue here.

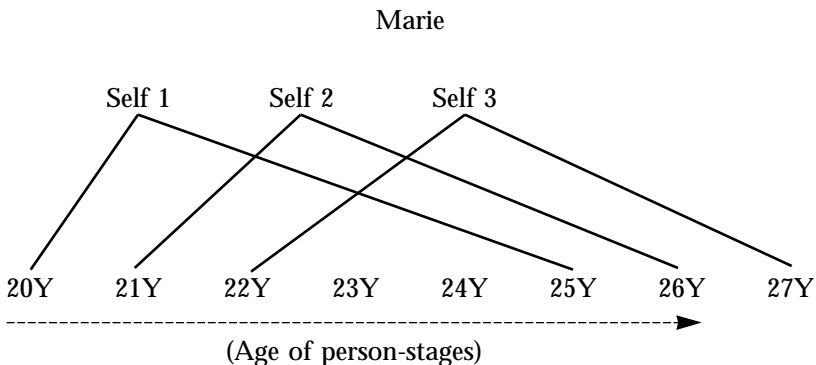


Figure 1³⁸

Furthermore, the various selves involved share a single body, and in order to carry out their long-range plans, they must interact and cooperate with one another:

But this means that [selves] will overlap with each other; they will stand to each other and the person much as strands of a rope stand to each other and the rope. Though we can recognize the overlapping strands as entities, the most salient entity is the rope itself. So, too, the most salient entity is the person, even if we can recognize the overlapping [selves] that make up the person ... These appear to be reasons for preserving the normal assumption that it is persons that are agents.⁴⁰

In other words, even though we can identify selves as metaphysical units that may serve some purpose in both questions of personal identity and even prudential reasoning, when we try to target them as moral agents we run into serious indeterminacy and confusion that threatens to undermine our general conception of agency. As a result, we should continue to target persons as moral agents for practical reasons.

Let me begin by mentioning that the kind of indeterminacy Brink has in mind would not be an issue in cases where there is a radical break in connectedness, e.g., in a person who underwent sudden memory loss with significant accompanying changes in character, beliefs, and intentions. In such cases, there would be no overlap and thus no indeterminacy about whose practical deliberation was at issue. But I have maintained all along that it is the *fact* of connectedness/disconnectedness – not the rate at which it takes place – that matters in the identification of selves, and in nearly all ordinary cases a loss of connectedness is indeed gradual, which would seem to imply the existence of the overlapping selves Brink worries about. So indeed Brink's objection apparently poses a serious challenge to my attempt to posit *ordinary* cases of selves as moral agents. What is to be said in response?

Brink offers two worries, both due to the overlapping nature of selves: first, we cannot non-arbitrarily identify the deliberating agent at any given time, and second, because the overlapping selves must cooperate anyway in order to act, it seems arbitrary to target them, rather than persons, as moral agents. But both worries are based on a misunderstanding about the nature of selves and what it would mean for selves to be targeted as moral agents.

To see why, consider once more *Figure 1*. I had earlier suggested (following Brink) that 23Y was a member of three different overlapping selves/agents, and the objection was that there was no way to identify which one's practical deliberation was in question. But this way of putting the matter is in fact rather misleading. Again consider Marie at twenty-three. Who is the Moderate Reductionist moral agent at this point? There is only one, viz., the unit corresponding to 23Y's *present self*. Remember,

one's present self is made up of the parts of one's life with which one is, or will be, strongly psychologically connected. The twenty-three-year-old Marie is strongly connected to both 20Y and 27Y. So when we inquire about the agent doing the practical deliberation *at age twenty-three*, the answer is that the deliberator is simply the twenty-three-year-old Marie's present self, the agent unified *at this point* from ages twenty to twenty-seven (and perhaps beyond) by strong psychological connectedness.

Now Brink might respond that this answer merely moves the problem back a step. For now what are we to say about the identity of the deliberating agent when Marie is twenty-six, for example? This stage (26Y) is no longer strongly connected with 20Y, as 23Y was, yet it remains strongly connected itself with 23Y and will be (let us say) strongly connected to stages in the future extending beyond the domain of 23Y (e.g., 30Y). So is 26Y the same agent as 23Y or not? If so, then strong connectedness cannot be sufficient for unifying agents (given that 26Y is strongly connected to stages that 23Y is not (e.g., 30Y), and it is not strongly connected to stages that 23Y is (e.g., 20Y)). If not, then the two agents must overlap, leaving us with the same problem as before: who is the practical deliberator in question at each stage? In other words, if 23Y's present self is one agent and 26Y's present self is another, yet they each include the other as stages within the scope of their own agency, then how can we say that the agent in charge of practical deliberation when Marie is twenty-three is really 23Y's present self and not 26Y's? Similarly, why assume that the deliberator when Marie is twenty-six is 26Y's present self and not still 23Y's? In fact, we have no non-arbitrary way of saying who the agent is in each case. So either psychological connectedness is not what matters in a determination of moral agency, or we are left with the confusions and indeterminacies described by Brink.

But the question, "Is 26Y the same agent as 23Y?" involves a false assumption, viz., that the *diachronic* identity of agents/selves is always a determinate matter. But it is not. Because the unity-relation for Moderate Reductionist agents/selves is psychological connectedness – a scalar relation – it *cannot* serve as a criterion for identity across time, as I have stated previously, given that it is not a transitive relation. X may be strongly connected to Y, and Y may be strongly connected to Z, but that does not at all mean that X will be strongly connected to Z. So when discussing the practical deliberation of 23Y, we may say that this is the practical deliberation of *an* agent, one constituted by the strongly-connected stages 20Y through 27Y. And when we discuss the practical deliberation of 26Y, we may say that this is the practical deliberation of an agent, one constituted by the strongly-connected stages 21Y through 30Y. Is this latter agent identical to or distinct from the former agent? This is an empty question, for there simply is no fact of the matter regarding their identity. So yes, there are indeterminacies at work here. We may describe

the various similarities and differences involved between the two stages (23Y and 26Y) in detail, but this is the most we can say about the identity of the various deliberating agents.

Indeed, though, this is all we *need* say. After all, what does it matter whose deliberation is in question? As Brink himself admits, we can easily identify the *reasons* for action of the entities at each stage.⁴¹ Why then must the identity of the “individual” deliberating agents be at all significant? Consider, for example, a somewhat analogous case, that of a club, call it Knights of the Square Table (KST) at time t_1 .⁴² Together, its members deliberate and vote on various actions that will represent their present interests (and the interests they foresee having) in the community. When they act, they act as a single (corporate) agent. Over the course of a few years, some members leave and are replaced (shifting the balance, say, of priorities for the club), and perhaps they change their name (now the Knights of the Oval Table (KOT) at time t_2). Some of the members of the club remain at t_2 , though, as do some of the club’s priorities. There remains, then, some “connectedness” between the club at t_1 and the club at t_2 . Are the clubs at t_1 and t_2 stages of the same agent? If not, are there two overlapping agents from t_1 to t_2 , then, the agents KST and KOT? If this is the case, then which one is the deliberating agent at t_1 ? At t_2 ?

Who cares? It seems obvious when we consider corporate agents such as this club that questions about the identity of the agents and practical deliberators at various times are completely unimportant. What matters instead is that there is an agent involved, both in deliberation and action, at each of the times in question. At t_1 there is an agent with identifiable interests and reasons for action, perhaps stemming from its past and reaching into (what it foresees as) its future, and at t_2 there are also identifiable interests and reasons for action. We can describe the similarities and differences between the clubs at t_1 and t_2 in detail, but there simply is no fact of the matter about their identity or about which one is the “agent in charge” at any given time. These questions are empty and insignificant.

This emptiness and insignificance also applies, I suggest, to similar questions regarding Moderate Reductionist selves/agents. One person-stage may very well be part of two selves/agents. What *matters*, though, is simply that there are, at that time, identifiable interests (stemming, say, from past stages with which it is presently strongly connected) and reasons for action (which project it into a future with which it intends to maintain strong connectedness). If these are the interests and reasons for action of two agents, then we may consider the two agents to be acting as *one* agent, much as we do with married couples or business partners who act as one agent where their interests coincide. Where interests and reasons for action coincide among two or more agents, the identity of the individual agents or an identification of who it is that is the primary practical deliberator is

irrelevant. The only thing that matters is that there is *an* agent at work at any given time. Consequently, Brink's worry is groundless.

What I have said thus far, however, does not rule out persons (unified by psychological continuity) as moral agents; I have merely defended the idea of selves as agents *as well*. But here is where the arguments of the first part of the paper enter in again, for there are practical reasons *not* to target persons as moral agents. What we have been focusing on recently have been cases in which selves overlap. As I have argued, these are cases that Moderate Reductionism can handle. But when we consider cases in which there are clearly two distinct selves at work, Moderate Reductionism becomes far more plausible than Conservative Reductionism.

Consider one last time my Revised Combined Spectrum case and the entities at the end points: I am at the near end; a Garbo Replica (who is continuous with me) is at the far end. Let us review, then, the conclusions we came to about this case. First, on a Psychological Criterion of personal identity, I am identical with the Garbo Replica. But in so far as personal identity is not the relation that matters in terms of survival and prudential reasoning about the future, we must focus instead on the relations that do matter, viz., psychological continuity and connectedness. Of these two, psychological connectedness is the relation that matters here, and it matters far more than continuity. Furthermore, the suggestion was made that, even though the Garbo Replica is continuous with me, she is enough like another person that she should be treated as such for moral purposes. The Garbo Replica and I are, therefore, to be treated as two distinct moral agents.

Conservative Reductionism, however, must maintain that the Garbo Replica and I are temporal parts of the same, overall agent, the single agent unified by psychological continuity, and this is true *even if* connectedness is the relation that matters for survival and prudential reasoning. But now the problem becomes clear, for this final fall-back position would create a serious tension, if not a complete schizophrenic breakdown, between the types of reasoning required of moral agents. Part of what it means for me to be an agent (as discussed earlier) is that I have interests which I am concerned to advance and which give me reason(s) to act. But which relation is to ground my reasoning about the future satisfaction of these interests in any given case? Indeed, which relation is to ground my reasoning about what interests I ought to cultivate? Who I am for moral purposes will differ from who I am for prudential purposes. If connectedness grounds my reasoning about my own survival – with its concomitant prudential interests – and continuity grounds my reasoning about moral interests, certain irresolvable conflicts will arise.

An example will serve to illustrate the point. Suppose that in the Revised Combined Spectrum case I am about to take the transforming pill, knowing full well what will happen during the next twenty-four hours. What

do I have reason to do? In my capacity as a prudential agent (unified by strong psychological connectedness), I will have reason to despair, knowing that what is about to happen to me is just as bad as ordinary death. So I may want to bid a tearful goodbye to my wife and children, make a last declaration, etc. But in my capacity as a moral agent (unified by psychological continuity), I will have reason to do no such thing. After all, as the agent who made the original promise to my wife to love and honor her, "I" will still be around to fulfill those moral obligations. So what reason would I have to bid her a tearful goodbye if I will continue to exist as her husband beyond the next twenty-four hours (despite undergoing some rather radical physical and psychological changes)? But which reasons for action should I honor? There would seem to be no way out of the dilemma.

And this would only be the tip of the iceberg. Suppose I had earlier committed some crime that I knew would be found out in forty-eight hours. As a prudential agent, I would have reason to be relieved (perhaps), knowing that I will not have to endure any time in prison. As a moral agent, however, I would have reason to be anything but relieved, knowing full well that "I" *qua* the Garbo Replica would be held morally responsible for the deed. Which reasons are to guide me? And which "me" is to be guided, the prudential agent or the moral agent? If all of this is beginning to sound hopelessly confusing, that is the point. And even when we move from the more contrived Revised Combined Spectrum case to ordinary cases, the problems would remain, for there would always remain a split between one's reasons regarding one's prudential agency and one's reasons regarding one's moral agency as long as a different relation were targeted as grounding each. This last attempt to rescue Conservative Reductionism fails.

5. Conclusion

While it may sound as if I have been beating an obviously dead horse, these latter remarks actually serve to buttress the argument of the present work as a whole. What I suggested early on was that it simply makes intuitive sense to target the same metaphysical units for both prudential and moral agency, but at the time this was nothing more than a suggestion. All I had established to that point was that there are solid metaphysical reasons for selves to be thought of as the units that matter for questions of survival and prudential reasoning. But now we are finally in a position to make a much stronger set of claims. First, the *practical* reasons offered against selves as prudential units miss the mark. Indeed, there remain both good metaphysical and good practical reasons to do so. Second, neither Extreme nor Conservative Reductionism offers a viable

alternative to Moderate Reductionism's emphasis on selves as *moral* units. On the one hand, Extreme Reductionist agents cannot, it seems, be agents at all; they exist for too brief a time to have either interests or reasons for action. On the other hand, Conservative Reductionist agents would have *plenty* of reasons for action – too many, in fact. They would be schizophrenic entities, frozen into *inaction* by their conflicting moral and prudential concerns.

It is, in fact, this last reason that undermines any attempt to target entities other than selves as moral units, for the point should now be clear that the relation serving to ground prudential reasoning should also be the relation serving to ground moral agency. To do otherwise (by taking either the Extreme Reductionist or Conservative Reductionist routes) is to split one reasoner into two, leading to hopeless confusion and unresolvable conflicts. Consequently, if we are to be Reductionists about personal identity, we ought to target selves as both our prudential and moral units. I set aside for now the many implications of such a view for ethical theory.⁴³

Department of Philosophy
University of California, Riverside

NOTES

¹ For comments on persons as possible moral units, see *Reasons & Persons* (Oxford: Clarendon Press, 1984), p. 322: "We may even think that only the killing of persons is wrong." For comments on selves as possible moral units, see *The Nineteenth Century Russian case*, pp. 327–9. And for comments suggesting momentary states of experience as the moral units, see p. 341: "It becomes more plausible, when thinking morally, to focus less upon the person, the subject of experiences, and instead to focus more upon the experiences themselves." (In what follows, when referring to *Reasons & Persons*, I will simply mark the referenced page number from this book in parentheses in the text.)

² Keep in mind that Parfit later argues that identity is simply not what matters, so this criterion basically ends up being unnecessary when it comes to matters requiring an account of what *does* matter in questions of survival and identity.

³ See *Reasons & Persons*, Section 82, for his treatment of this topic.

⁴ Briefly, the argument runs as follows: if we claim (as seems natural) that I would survive in the case in which one hemisphere of my brain were successfully transplanted into only *one* brother, then why would we claim I cease to exist when one of my hemispheres is successfully transplanted into *each* brother? How could a double success constitute a failure? This question makes the first possibility hard to believe. As for the second and third possibilities, what could possibly make it the case that I have survived as one brother and not the other? Both resulting people would be exactly similar to me psychologically (and, in all relevant respects, physically). Finally, to suggest that I have somehow survived as both brothers would greatly distort our concept of personhood, for the brothers might go their separate ways, have wildly varying experiences, etc. To say that both of these spatially distinct entities were still me would be to stretch our concept of personhood into incoherence.

⁵ Mark Johnston objects to this inference, arguing instead that personal identity *is* what matters and Relation R only comes to the fore in the odd science fiction cases, cases which

rarely, if ever, occur. See his "Reasons and Reductionism," *The Philosophical Review* (July 1992), esp. pp. 613–17. However, whether it is Johnston or Parfit who is right here is irrelevant for my purposes (although I must admit to agreeing with Parfit on this point). My aim in the present work is not to argue for the correctness of Parfit's picture; rather, I wish to argue that *if* Parfit is right about the metaphysics of identity, then he is at worst wrong, and at best misleading, about some important implications of that view for ethical theory.

⁶ The word "I," then, we would use to refer only to this present self.

⁷ I say more about what this attitude involves shortly.

⁸ Although he emphasizes elsewhere that talk of successive selves is a mere *façon de parler*. See his Postscript to a reprinted version of "Later Selves and Moral Principles," in Ted Honderich and Myles Burnyeat (eds), *Philosophy As It Is* (New York: Penguin Books, 1979), p. 211.

⁹ See Parfit's "Later Selves and Moral Principles," in A. Montefiore (ed.), *Philosophy and Personal Relations* (London: Routledge & Kegan Paul, 1973), pp. 139–42.

¹⁰ See Parfit's "On 'The Importance of Self-Identity,'" *Journal of Philosophy* 68 (1971): 683–90 for a discussion of this sort of indifference. See also my "Theoretical Persons and Practical Agents," *Philosophy & Public Affairs* 24 (1996): 318–32, esp. p. 329.

¹¹ Although, oddly enough, he claims outright on p. 206 that "[o]f these two general relations, connectedness is more important both in theory and in practice." There is no argument provided in support of this claim, however. My aim here is to provide one.

¹² This sort of indifference, however, does not at all imply a lack of any concern whatsoever. What I have in mind here is that the relation between selves is like the relation between (what are normally termed) persons. What is missing is merely the sort of *special* concern we generally find in my relation to various stages of my (present) self.

¹³ It is fairly clear that Parfit himself would like to appeal to this conclusion as well, given his arguments regarding rationality and morality (Chapters 14 and 15). But because he also seems committed to the view that the importance of continuity and connectedness is roughly equal, I fail to see how the targetting of selves as the morally significant units is warranted in *Reasons and Persons*. Of course, I see the argument given herein as one way for him to support his arguments regarding rationality, but I believe it has significantly non-Parfitian ramifications for morality, as I intend to argue elsewhere.

¹⁴ Assuming, of course, that what matters in morality is independent of what is involved in prudential rationality, a fairly controversial assumption in its own right. If, on the other hand, moral principles are parasitic on principles of prudential rationality (e.g., the factors relevant for moral deliberation are derived from a form of foundational ethical egoism), and the metaphysical unity-relation that matters in terms of justification for intrapersonal deliberation is psychological connectedness, then the self (the significant unit of prudential rationality) would obviously also be the morally significant unit. But my position in this paper is that the self ought to be viewed as the morally significant unit *regardless* of the foundational ethical theory at work (given the correctness of Parfit's basic picture), so I will assume for the sake of argument that my hypothetical objectors in what follows are non-egoists (even though one of my *actual* objectors in what follows, David Brink, is an admitted rational egoist; nevertheless, this fact is insignificant in terms of the arguments about agency he presents).

¹⁵ Of course, whether or not teletransportation is indeed a form of travel, rather than simply a clone machine, is part of what is at issue in the case.

¹⁶ Parfit discusses both of these cases throughout Part Three of *Reasons and Persons*.

¹⁷ See his "Comments," *Ethics* (July 1986): 838. I am skeptical, however, that most of us *would* actually side with Backup once we realized that Backup remembers intending to get the goodies from the crime and also remembers enjoying the goodies. But perhaps those of us who think this way have already been "tainted" by our philosophical ruminations on personal identity. Still, for those who would side with Backup, Parfit is probably right to call

them Non-Reductionists, and these might be people who would then be led to the Extreme Reductionist position, once their original position were undermined by the Parfitian line.

¹⁸ *Ibid.*, p. 839.

¹⁹ I attempt to show its problems in "Disintegrated Persons and Distributive Principles," unpublished paper, 1999.

²⁰ Indeed, other arguments have been given in its support. Parfit discusses a different argument for such a position in his "Comments," pp. 839–42, as well as in *Reasons and Persons*, pp. 342–5. In addition, Brink offers a possible argument for Extreme Reductionism by focusing on the complaints I-now might have with regard to undergoing a burden now for the sake of a compensatory benefit to me-later. See his "Rational Egoism and the Separateness of Persons," in Jonathan Dancy (ed.), *Reading Parfit* (Oxford: Blackwell, 1997), pp. 110–11.

²¹ This view would, of course, have enormous implications for both prudential rationality and morality. For one thing, it would have to be prudentially irrational to sacrifice *anything* now for the sake of compensatory benefits to be distributed at any time other than the present, i.e., simultaneous compensation would be the only true form of compensation. See both Brink (p. 111) and Parfit (*Reasons and Persons*, pp. 342–3). In addition, I argue elsewhere that this is the only view of moral agency that provides the foundational support necessary for utilitarianism to be viable. See my "Utilitarianism and Personal Identity," *The Journal of Value Inquiry* 33 (June 1999): 183–99.

²² Brink, p. 112.

²³ *Ibid.*

²⁴ See Bernard Williams, "Persons, Character and Morality," in Amelie Oksenberg Rorty (ed.), *The Identities of Persons* (Berkeley: University of California Press, 1976), esp. pp. 206–7. See also Brink, pp. 112–13.

²⁵ *Ibid.*, p. 209.

²⁶ *Ibid.*, p. 206; emphasis in original.

²⁷ *Ibid.*, p. 209.

²⁸ Thus, I think Brink's claim that the momentary experiencer "will not persist long enough to perform actions or receive the benefits of actions" and so "cannot have reasons for action" (p. 112) is too quick. Even if I am really just a momentary experiencer, I may *conceive* myself as being of greater endurance (I am deluded, say, about my true nature) and so may take myself to have reasons for action. My reasons, after all, may be bad or irrational, but they are still reasons.

²⁹ Brink, p. 119.

³⁰ *Ibid.*, p. 120.

³¹ *Ibid.*

³² See my "Theoretical Persons and Practical Agents," esp. pp. 328–31. What I call there the "S-View" is more or less what I mean by Moderate Reductionism here.

³³ See her "Personal Identity and the Unity of Agency: A Kantian Response to Parfit," *Philosophy & Public Affairs* 18 (1989): 123 and *passim*. I should stress, though, that Korsgaard thinks the concept of authorial connectedness actually *undermines* the Parfitian project. In "Theoretical Persons and Practical Agents" I argue otherwise along lines similar to above.

³⁴ Parfit discusses this sort of identification in "On 'The Importance of Self-Identity,'" pp. 683–90, as do I in "Theoretical Persons and Practical Agents," pp. 329–31. It is important to point out, though, that my being able to identify with some past stage of myself is not constitutive of there being strong connectedness between us; rather, this attitude merely *reflects* the pre-existing connectedness that obtains.

³⁵ See "Theoretical Persons and Practical Agents," p. 330 for a more detailed discussion of this concept.

³⁶ I mention here just four: Brink, pp. 113–16; Williams, pp. 202–10; Korsgaard (throughout); and Susan Wolf, "Self-Interest and Interest in Selves," *Ethics* 96 (1986): 704–20.

³⁷ I believe the other objections regarding so-called pragmatic problems with targetting units other than persons as moral agents have been adequately answered elsewhere. For instance, Parfit himself has addressed the worries offered by Williams regarding the non-scalar dimension of promises in *Reasons and Persons*, pp. 326–9. As for Wolf's complaints about the practical problems associated with targetting just Relation R as the relation of moral importance in the fission case, Parfit's "Comments" seem to provide an adequate reply. And I have attempted to deflect Korsgaard's Kantian objections to the possibility of selves as moral units in my "Theoretical Persons and Practical Agents." Brink's remarks, however, are both quite recent (and thus reflect awareness of Parfit's own responses to the objections of Wolf and Williams) and more encompassing than these others (in that he wants to identify at a higher level of abstraction the complaint common to all these objections), and so a defense against his objection would, I believe, provide a general defense against objections of this kind.

³⁸ Brink sets out a similar diagram as Figure 6.1, p. 113.

³⁹ *Ibid.*, p. 114.

⁴⁰ *Ibid.*, pp. 114–15.

⁴¹ *Ibid.*, p. 114.

⁴² Parfit is also fond of comparing persons to both clubs and nations. See, e.g., pp. 242–3, 277–8.

⁴³ I am indebted to Eric Cave, Terry Horgan, James Hudson, Diane Jeske, Alan Nelson, Derek Parfit, Martin Schwab, Mark Timmons, and Gary Watson for their extremely helpful remarks on earlier drafts of this paper. I am also grateful for the insightful and generous comments on prior versions of this paper by audience members at the 1997 Mid-South Philosophy Conference, the 1998 Southern Society for Philosophy and Psychology Conference, and a 1999 University of Mississippi Philosophy Colloquium.