Over the last forty years, as the rest of analytic philosophy of mind has taken an empirical turn, personal identity theory has remained steadfastly preoccupied with metaphysics and non-empirical normative concerns. In the present paper, I want to suggest a way for personal identity theorists to give a higher priority than they have to empirical concerns and also provide them with a motive for doing so. The empirical concerns that I have in mind are empirical versions of two theses: that personal identity is not what matters primarily in survival; and that personal identity based egoistic values are not fundamental, but derived. Like any empirical theses, these will live or die on their ability to explain. They may also suggest future lines of research and have some practical significance.

To get a feel for what I have in mind, consider an analogy. Suppose that the only behavior that someone had ever perceived as fatally risky, and also the only thing that he had ever feared, was riding on roller coasters. It might seem plausible to suppose that *fundamentally* this person is afraid of roller coasters. However, another possibility is that fundamentally he is afraid of death and only derivatively afraid of roller coasters. How could we discover which, if either, of these were true?

Suppose that we were to learn that in addition to being afraid of riding roller coasters this person would also be afraid of other behaviors which he perceived as fatally risky.
Then, by appeal to his fear of death, we could explain—in a unified way—not only his fear of roller coasters, but fears that he would have if he were faced with other apparently fatally risky behaviors. In addition, we might be able to explain certain aspects, otherwise inexplicable, of his fear of roller coasters, such as why some parts of the ride—the ones in which he takes himself to be in greater danger of dying—are scarier to him than other parts. By contrast, there would be no explanatory advantage to the thesis that fundamentally he is afraid of roller coasters.

This person’s knowing that fundamentally he is afraid of death and only derivatively afraid of roller coasters could lead him to refocus his self-understanding around his most basic fears. This deeper self-understanding, valuable in itself, might, then, set the stage for significant changes in his attitudes and behavior. For instance, knowing this about himself, it might then be easier for him to empathize with people who are in life-threatening situations and he might acquire a stronger motive than he would otherwise have had to come to their aid.

Could anything like this happen in connection with the two theses that I mentioned: that personal identity is not what matters primarily in survival; and that personal identity based, egoistic values are derived? I think so. Suppose, for instance, that some fission examples were to reveal that in certain situations many of us, on egoistic grounds, would actually prefer cessation by fission to options in which we could have persisted. Suppose also, as I have argued elsewhere, that in certain situations in which our prospects, without undergoing fission, were quite bright—as bright or brighter than the prospects of any of us here today have ever been—many of us would still have this preference.1 The truth of these suppositions would strongly suggest that when it comes

Proto-egoism [60] page 2
to the future, it is not personal identity, but rather other values that are on the ground floor of our motivational systems—that is, the systems of those of us who have such preferences. Knowing this about someone might motivate a more refined, empirical exploration of his or her values. Knowing this about oneself might diminish the grip that the notion of personal identity has on one’s psyche, thus changing one’s attitudes and behavior toward a variety of issues, including one’s own impending death.

But should philosophers be in the business of explaining values? Isn’t that a job for psychologists? It is a job for psychologists. But, on this issue, philosophers can contribute importantly as psychologists in a way that it’s unlikely that psychologists proper will themselves contribute. The reason for this is that it is only by considering hypothetical examples that plausible candidates for what matters primarily in survival other than identity can be isolated from each other so as to generate evidence that personal identity is less important than we may have thought that it was. It seems very unlikely that psychologists proper, at least any time soon, are going to consider such examples.

In the personal identity literature, the theses that what matters primarily in survival is not personal identity and that personal identity based, egoistic values are derived have been closely linked—so closely, that it is possible to imagine that the two are the same thesis. But, neither thesis has been formulated precisely. So, it’s hard to tell what the relationship is between them. As I shall understand them, the empirical thesis that personal identity is not what matters primarily in survival implies the empirical thesis that at least future-oriented, personal identity based, egoistic values are derived. I shall return to the question of how the latter thesis might be understand more precisely. First,
though, I want to provide a motive for understanding it more precisely. This will involve suggesting why, and how, its being true would matter. It will also involve disentangling both theses from their involvement in the literature with their normative counterparts. The empirical thesis that future-oriented, personal identity based, egoistic values are derived is worth understanding more precisely since it is more easily defended than its normative counterpart, yet still has important consequences.

Since 1970 various philosophers—particularly Derek Parfit, but also Sydney Shoemaker, Robert Nozick, John Perry, and others—have argued that personal identity is not what matters primarily in survival. In my view, their arguments and the theories which they have proposed to attend them have been among the most important developments in personal identity theory in almost two hundred years. Yet, since the mid-70s both their arguments and their theories have met with staunch and growing resistance. The critics have included David Lewis, Peter Unger, Ernest Sosa, Susan Wolf, Christine Korsgaard, Simon Blackburn, John McDowell, Mark Johnston, Lynn Baker, and David Brink, among others.

What could explain the widespread resistance, by so many good philosophers, to what can seem to be such a genuinely progressive development? One possibility, of course, is that the development has not been progressive, but misguided. I reject this possibility. Another is that the critics have been fundamentally mistaken. I also reject this possibility. A final possibility is that in the debate over what matters primarily in survival, more than one important thesis has been under discussion, and that while the critics have successfully undermined some of these theses they have failed to undermine others. In my view, this latter possibility captures what has actually

Proto-egoism [60] page 4
happened.

For present purposes, the important distinction among different versions of the thesis that personal identity is not what matters primarily in survival—hereafter, the radical thesis—is that between a descriptive and a normative version. The descriptive version is that as a matter of fact what matters primarily in survival to many people whose values are reasonable is not that there be someone in the future who is the same person as oneself, but that one be related in certain ways to someone in the future—ways which do not necessarily insure that this person is oneself. The normative version, on the other hand, is not about what as a matter of fact does matter primarily in survival to many or, for that matter, about what does matter to anyone, but about what should matter primarily to everyone. It is the thesis that what should matter primarily in survival is not that there be someone in the future who is the same person as one is now, but that one be related in certain ways to someone in the future—again, ways which do not necessarily insure that this person is oneself.

As anyone familiar with the personal identity literature will recognize, it has been the normative version of the radical thesis that has been at the center of controversy, while the descriptive version has been relatively ignored. There are, I think, two fairly obvious reasons for this: first, to most theorists the normative version has seemed the bolder and more interesting of the two theses; and, second, Parfit, in particular, has argued—in two main ways—very provocatively for the normative version, thus luring many theorists into discussing it. Parfit has argued for, first, on the basis of a certain sort of metaphysical reductionism, which I shall call austere reductionism and, second, on the basis of fission examples.

Proto-egoism [60] page 5
Of these two sorts of arguments, Parfit’s arguments based on austere reductionism have drawn most of the critical fire. Since some of his critics seem to have assumed that if these arguments are defeated, then the radical thesis itself has been defeated, it is important to recognize that the arguments based on austere reductionism and those based on fission examples do not stand or fall together. The easiest way to see this is to suppose that people have something like Cartesian egos, but ones that split amoeba-like under the pressure of fission. If that were the case, then austere reductionism would be false, but most of Parfit’s fission example arguments—and the fission example arguments of others—for the radical thesis would be relatively unaffected.

These two ways of arguing for the radical thesis differ from each other in another important way. The arguments based on austere reductionism always, or almost always, purport to be rationally coercive—specifically, to show something about what people are required to value, while typically those based on fission examples purport merely to reveal what people already do value. For present purposes, this difference is crucial, first, because arguments based on austere reductionism tend to be more vulnerable to objections than many of those based on fission examples and, second, because arguments based on fission examples more readily lend themselves to establishing empirical results.

In arguing for a deflationary reevaluation of the importance of personal identity, Parfit has not only highlighted the bolder normative version of the radical thesis, rather than the more cautious descriptive version, and argued for it importantly on the basis of

Proto-egoism [60] page 6
austere reductionism, but, in his most visible formulations of the thesis, has characterized what matters primarily in survival in an overly simple and overly quantitative way; for instance, by saying that what matters primarily is relation-R and that one is R-related to someone in the future just if there is "over any day at least half the number of direct [psychological] connections that hold, over every day, in the lives of nearly every actual person." That he has done all of this so provocatively has contributed greatly to the interest that others have taken in the radical thesis. But the provocation has been due importantly to his having argued brilliantly for a extreme version of the thesis.

Consider, for instance, Parfit's notorious “branch line” case. Contrary to what he has argued largely on the basis of austere reductionism should matter to us in that situation, most people who project themselves into the position of the person who enters the teletransporter find that they care more about their descendent who remains on Earth than about their Martian replica. Moreover, as Mark Johnston has pointed out, most people would care more about their descendant on Earth even if, on a quantitative construal of relation-R, they were less psychologically connected to him than to their Martian replica.

Or, imagine a variation on the branch line case in which the person who enters the teletransporter on Earth:
· is uncertain whether he will be dematerialized or walk out of the teletransporter;
· believes that if he is not dematerialized, then he will be identical with the person who walks out of the teletransporter;
· believes that if he is dematerialized and moments later a replica of himself is...
produced on Mars, then he will be identical with his descendant on Mars;
and believes that the mission on Mars will be very dangerous.

On egoistic grounds, what would a person with such beliefs hope, as he enters the
teletransporter? If he is like most of us, probably he would hope that on pushing the
button to activate the teletransporter, he will *not* be dematerialized but, instead, walk
out of the teletransportation booth unharmed, leaving his replica to preform the
dangerous mission on Mars. This is also what his friends would wish on his behalf. vi

What such considerations strongly suggest, I think, is that in the case of many of us
whose values are reasonable, relation-R is not what matters primarily in survival.
Rather, what matters primarily is *either* identity *or* some relation more closely tied than
is relation-R to identity, but yet still weaker than identity. Perhaps we could be
converted to the view that relation-R should be what matters primarily, say, by being
persuaded to accept austere metaphysical reductionism, but this seems unlikely.
Currently many good theorists remain unpersuaded. Yet several fission examples other
than the branch line case strongly suggest that identity is not what matters primarily.

Where, then, does this leave us? Where it seems to leaves us is with the realization
that while some fission examples seem to show that identity is not what matters
primarily in survival, others, such as the branch line case, seem to show either that
identity is what matters or that what matters, while not identity, is richer than relation-R.
In my view, the latter option is the right one. In other words, what we should conclude is
that in the case of many of us whose values are reasonable, what matters primarily in
survival is not that there be someone in the future who is the same person as one is
now, but that one be related in certain other ways to someone in the future—ways which
are richer than relation-R in being more closely related to personal identity, but which are still weaker than personal identity. If that were true, then future-oriented, personal identity based, egoistic values would be derived.

In sum, in my view, what explains the resistance by so many good philosophers to the radical thesis is five things: first, that the thesis itself is radical in that it flies in the face of common sense; second, that it is radical in that it threatens well-entrenched normative theories; third, that the debate over it has been almost exclusively over normative versions of it; fourth, that the debate over it has highlighted arguments based on austere reductionism; and, fifth, that some theorists seem to have assumed that if the arguments for it which are based on austere reductionism go down, then the radical thesis goes down with them. But except for the first two of these—conflicting with common sense and threatening well-entrenched normative theories—these sources of resistance would be misdirected against the descriptive version of the radical thesis.

III

Suppose, then, for the sake of argument, that the descriptive version of the radical thesis were true and, hence, that future-oriented, personal identity based, egoistic values were derived. Would this version of the radical thesis’ being true make any difference? The obvious answer, I think, is that its being true would seem to make the following difference: that in certain kinds of choice situations, its being true would diminish the deliberative weight of the consideration: “But that person in the future will (or will not) be me.” I shall call this consideration, the me-consideration. By deliberative weight, I mean how much, in a person’s trying to figure out what to do, a consideration counts for or against her electing a certain option. By something’s diminishing the

Proto-egoism [60]  page 9
deliberative weight of a consideration, I mean its causing that consideration to count less in a person’s decision whether to elect an option.

In what kinds of choice situations would the descriptive version of the radical thesis being true diminish the deliberative weight of the me-consideration? Basically, in two kinds: first, in certain hypothetical situations involving fission and, second, in certain other kinds of hypothetical and actual situations in which one believed that one’s future self would become disconnected dramatically to one’s present self. In the first sort of (fission) situation, one might feel that what matters primarily is not necessarily that some person in the future who will benefit from one’s choice is oneself, but that one be related to such a person in ways that matter primarily to one in survival. In the second sort of (non-fission) situation, one might feel that what happens to one’s future self is of no more concern than what happens to anyone else.

This latter sort of situation might arise if one knew, or thought that one knew, that in the future one would become amnesiac with respect to one’s present self. It might also occur if one knew, or thought that one knew, that one would adopt values that one presently finds abhorrent. In such cases, the special concern that normally one would have for one’s future self might be blocked. Psychologically something like this happens in certain sorts of psycho-pathology, such as depersonalization or alien limb syndrome, in which the sympathy that normally one would have for oneself or for parts of oneself is blocked. At the very least, in such cases one might experience diminished self-concern.

If, as I believe, these seeming consequences of some properly qualified descriptive

Proto-egoism [60] page 10
version of the radical thesis are in fact its consequences and some such view is true, then in spite of the criticisms that have been urged against certain ways in which the radical thesis has often been formulated and defended, the thesis is not only correct but, like its normative cousin, still has fangs.

III

Recently Mark Johnston has tried to defang the radical thesis by arguing that we should “quarantine” the responses that many people have had to fission and other hypothetical examples. His point is that even though in certain bizarre cases some of us would be willing reasonably to extend to others the pattern of special concern that ordinarily we reserve only for ourselves, identity is still what matters primarily in survival.

Johnston admits that in some fission cases it would be reasonable to care about each of one’s fission-descendants as if each were oneself. He says, for instance, that even though he does not believe that he would “determinately survive” a familiar sort of fission procedure in which his cerebral hemispheres were divided and both were transplanted into different receptacle bodies, but does believe that he would survive such a procedure in which just one of his hemisphere were transplanted, he would still not “make a significant sacrifice” to ensure that just one was transplanted. However, unlike radicals who have concluded from such a consideration that identity is not what matters primarily in survival, Johnston says that the most that it would be reasonable to conclude is that in cases like this “(neurally based) R, and not identity, is the relation in terms of which one should extend one’s special concern.” One would not be warranted, he claims, in concluding “that quite generally it is (neurally based) R that matters.”

Proto-egoism [60] page 11
Why wouldn’t one be warranted? According to him in such fission examples, at least two presuppositions of our special concern are violated: first, that it is always a determinate matter whether one is identical with some given future person; and, second, that at most one future person will continue one’s mental (and physical) life. When such presuppositions are violated, he says, future-directed concern neither determinately applies nor determinately fails to apply. Hence, in such cases, it is reasonable to try to find a natural extension of future-directed concern. When a significant core of the relations which constitute identity in the determinate cases is present, “an appealing idea,” he says, is to extend one’s special concern to one’s fission descendants. In his view, the holding of neurally based psychological continuity and connectedness, a relation that in many fission examples holds twice over, constitutes such a significant core. But, he insists, this would be reasonable “not because identity is never what matters” but, rather, because “caring in this way represents a reasonable extension of self-concern in a bizarre case.”

Johnston says that in response to the bizarre cases, it would not be reasonable to move away from the view that identity is what matters primarily to the view that relation-R is what matters because such a move would be “much more radical” than his more conservative move to locally modified concern and is more than the indeterminate cases require. He concludes that “there is a false apparatus of generalization at the heart of Parfit’s arguments against identity-based concern.” The “relevant presupposition of self-concern is the holding of the determinate, ordinary fact of personal identity or difference.” Since this presupposition is almost always met, fission examples have “no effect beyond the imaginative fringe. What is not there in the fission case,” he says, “is almost always there. Identity is still almost always what matters.”
I agree with Johnston that it does not follow from the consideration of fission examples that identity does not matter primarily in survival. Nevertheless, there is, I think, a serious problem with his argument. The problem is with the way in which he characterizes the data that need to be explained. In his view, if we project ourselves into the role of the protagonists in the sorts of fission examples under consideration, then we are presented with a puzzle—whether to extend special concern to our fission-descendants. He claims that we can solve this puzzle reasonably either by extending our concern or by withholding it—that it is up to us.

But this characterization of the way in which many people respond to fission examples gets the phenomenology wrong. For many of us, in projecting ourselves into the place of the protagonists in some fission examples, it is not a question of whether to extend our special concern to our fission descendants; rather, our pattern of special concern already extends to our fission descendants. As a consequence extending our special concern is not something that in response to fission examples we decide to do, but something that in considering them, we discover that we have already done.

In other words, contrary to Johnston’s characterization, in the case of many of us, in contemplating certain fission examples there is no deliberative problem that needs to be solved. Rather, the examples reveal something about what we already value. What they reveal is that our special concern already extends to our fission descendants—that we already have a strong preference in the choice situations depicted in such examples, to adopt options that conflict with what philosophers traditionally have supposed is what matters primarily in survival.
What explains these revealed preferences? In Johnston’s account, so far as he goes, nothing explains them: we simply value one thing in one situation and another in a different situation; our preferences change willy nilly. By contrast, if what matters primarily in survival to those of us who feel special concern for our fission descendants were not personal identity, but something else—something that obtains both in normal cases of self-concern and in the hypothetical fission examples under consideration, then a unified explanation of our preferences could be provided.

Prima facie, then, there is reason to think that an empirical version of the theses that personal identity is not what matters primarily in survival and that future-oriented, personal identity based, egoistic values are derived are both true and consequential, hence worth further empirical investigation and elaboration. But for this to happen fruitfully, the theses have to be formulated more precisely. How should this be done? I want next to suggest a way that it might begin to be done in connection with the latter thesis: that future-oriented, personal identity based, egoistic values are derived.

IV

Values, in the sense in which I am using the term, are things that one wants. A person realizes her values if she gets what she wants. So, if a person wants to write a good book and she writes one, then she gets what she wants and thereby realizes that part of her values. Of course, a person might get some things that she wants—say, world peace—not by her doing anything in particular, but simply by something, such as world peace, happening.

By egoistic values, I mean those values which a person embraces because she thinks
that their realization will benefit herself primarily, rather than others. For instance, a person’s desire to be happy, not in order to benefit others, but to improve her own life, expresses an egoistic value that she has—her own happiness. Her desire that others be happy, not in order to benefit herself, but to improve their lives, expresses a value that she has that is not egoistic—the happiness of others.

By personal identity based, egoistic values, I mean those egoistic values which a person embraces at least partly because she thinks that their realization requires that she exist.

In the what-matters-primarily-in-survival literature, it has been future-oriented, personal identity based, egoistic values, rather than past- or present-oriented values, that have loomed large in the debate. By future-oriented, personal identity based, egoistic values, I mean those personal identity based, egoistic values which a person embraces at least partly because she thinks that their realization requires that she exists at some point in the future.

Finally, by proto-egoistic values, I mean those values, if any, from which egoistic values may be derived.

Future-oriented self-concern is the concern that one has for someone who will exist, or who one imagines will exist, in the future because one thinks that that person is, or would be, oneself; for instance, I care especially and in a distinctive way about what will happen tomorrow to that person who is (or will be) me because that person is (or will be) me.
Self-referential concern, on the other hand, is the concern that one has for persons or things because they stand in certain special relationships of ownership or appropriation to oneself. For instance, I care especially and in a distinctive way about what will happen tomorrow to my children because they are my children and to my car because it is my property. In short, self-concern, in the sense in which I am understanding it, is concern for what is (or was, or will be) oneself because it is oneself, while self-referential concern is concern for what is (or was, or will be) one’s own because it is one’s own.

Typically, future-oriented, personal identity based, egoistic values get expressed in self-concern, not in self-referential concern.

Against the backdrop of this stage-setting, what, then, does it mean for a person’s future-oriented, personal identity based, egoistic values to be derived? What it means is that a person embraces non-egoistic values which are such that were she to realize these, she would also realize her future-oriented, personal identity based, egoistic values. How this would work in practice depends, in part, on the metaphysics of persons.

Consider, first, how it would work on a three dimensional view of persons. Suppose, for instance:

- that what matters to me primarily in survival is being related to someone in the future in a certain way—say, by psychologically continuous with and connected to (R-related to) someone;

Proto-egoism [60] page 16
that being R-related to someone in the future does not necessarily ensure that I will be R-related to myself in the future;

that I egoistically value being R-related to myself in the future; and

that in the circumstances in which I find myself, the only way to be R-related to someone in the future is to be R-related to myself in the future.

In such a situation, my future-oriented, personal identity based, egoistic value of being R-related to myself in the future would be derived, rather than fundamental. It would be derived since I want to be R-related to someone in the future, but not necessarily to myself, and in the circumstances can accomplish this by being R-related to myself and only by being R-related to myself.

On a four dimensional view of persons, people who currently exist cannot, of course, exist wholly in the future. So, we would have to describe this same example differently. Rather than saying that one egoistically values being related to oneself in the future, we should say that one egoistically values being related to future person stages which are temporal parts of oneself.

To see what difference this makes, suppose, as before, that what matters to me primarily in survival is being R-related to someone in the future, but that I also egoistically value being R-related to myself in the future. Then, on a four dimensional view, my personal identity based, egoistic value of being R-related to myself in the future would be derived if two conditions are satisfied: first, it were not possible for any current person stage of mine to be R-related to anyone whose future person stages do not include my present person stage as a temporal part; and, second, that the reason for this is that the unity-relation among person stages has by stipulation been made to

Proto-egoism [60] page 17
match that of the independently determined R-relation (as David Lewis put it, “the I-relation is the R-relation”). In such circumstances, what would make the personal identity based, egoistic value of being R-related to myself in the future derived is that the unity relation—the I-relation—and the R-relation are the same and that the R-relation, as it were, leads the way, while the I-relation follows its lead.

However, if this is how the personal identity based, egoistic value of being R-related to myself in the future is derived, then the thesis that it is derived is not an empirical thesis. For it to be an empirical thesis, it cannot have been arranged by stipulation that the I-relation and the R-relation are the same. Suppose, then, that we do not stipulate that the I-relation and the R-relation are the same and that, as it happens, they sometimes diverge. In such a case, what it would mean for my future-oriented, personal identity based, egoistic value of being I-related to myself in the future to be derived is that there are no actual or possible cases in which my current person stage is I-related to some future person stage to which it is not also R-related.

So much, then, for what it means for future-oriented, personal identity based, egoistic values to be derived. What would it matter if they were? The most disruptive implication would be the weakening of the me-consideration in complicated ways that still have to be worked out. Although the bottom line has yet to be written, it seems clear that however it is written, if future-oriented, personal identity based, egoistic values were derived, their being derived would pay havoc with a great deal of what many normative theorists have wanted to say either about prudence and self-interest or about their importance. The heart of the disruption would not be that something other than personal identity, such as relation-R, should matter primarily in survival. Rather, it
would be that the personal identity, in the guise of the me-consideration, would have been revealed as less important than traditionally has been supposed. And it would have been revealed as less important not by normative, but empirical considerations; that is, not by coercive arguments about what we should value, such as, Relation-R, but mostly by examples which reveal that many of us already value personal identity less than we thought that we did and by considerations which may cause, but would not rationally require, many of us to value it even less. Of course, someone could always argue that although personal identity is as a matter of fact not as important a value as traditionally has been supposed, it should be as important. But the prospects of anyone’s arguing successfully for this normative thesis do not seem bright.*

**ENDNOTES**

* Thanks to Allen Stairs for helpful suggestions and to Reina Hayaki for a very useful comment on an earlier version of this paper.

________________________


iii. Parfit, Reasons and Persons, p. 222. Some radicals have suggested that the ways of being related to someone in the future that matter have to be expressed not merely quantitatively, but qualitatively, and that these ways are not the same for everyone, but vary from person to person. See, for instance, Nozick, Philosophical Explanations, p. 000. In a passage that occurs rather late in Parfit’s defense of the radical view in Reasons and Persons, he concedes that relation-R has to be characterized qualitatively, p. 000.


vi. Ibid.


viii. Mark Johnston has done this nicely in “Human Concerns without Superlative Selves,” pp. 156-58, in Dancy, Reading Parfit, pp. 149-77.

ix. David Lewis, op.cit.

Proto-egoism [60] page 20