

DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY


Personal Identity and the Unity of Agency: A Kantian Response to Parfit

Citation

Korsgaard, Christine. 1989. Personal identity and the unity of agency: A Kantian response to Parfit. Philosophy and Public Affairs 18, no. 2: 101-132.

Published version

http://www.jstor.org/stable/2265447

Link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:3219881

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFlOWI2M2EwYTg

Accessibility

https://accessibility.huit.harvard.edu/digital-accessibility-policy

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story

Personal Identity and the Unity of Agency:
A Kantian Response to Parfit[1]

Christine M. Korsgaard

I. *Introduction*

A person is both active and passive, both an agent and a subject of experiences. Utilitarian and Kantian moral philosophers, however, characteristically give a different emphasis to these two aspects of our nature. The Utilitarian emphasizes the passive side of our nature, our capacity to be pleased or satisfied, and is concerned with what happens to us. The Kantian emphasizes our agency, and is concerned with what we do. Alternatively, we may say that the Utilitarian focuses first on persons as objects of moral concern, and asks, "what should be done for them?", while the Kantian addresses the moral agent, who is asking, "what should I do?"

One might think that this can only be a difference of emphasis. Any acceptable moral philosophy must take both sides of our nature into account, and tell us both how people ought to be treated and what we ought to do. Yet the difference of emphasis can lead to substantive moral disagreement. Kantians believe in what are sometimes called "agent-centered" restrictions, obligations which are independent of the value of the outcomes they produce.[2] Even when thinking of persons as objects of moral concern, the Kantian is more likely to focus on agency. The question, "what should be done for them?" is answered, roughly, "they should be given the freedom to make their own choices, and to do things for themselves." Rawls believes that asking the agentless "what should be done for them?" leads to distortion in the Utilitarian view of moral and

political decision. The idea that burdens for some people can be justified simply by benefits to others:

> … arises from the conception of individuals as separate lines for the assignment of benefits, as isolated persons who stand as claimants on an administrative or benevolent largess. Occasionally persons do so stand to one another; but this is not the general case…[3]

When persons are viewed as agents who are making agreements with one another, this way of looking at their relations is not the natural one.

Of course the Utilitarian claims to take agency into account. He acknowledges that persons do not just want things to be done for them but to do things; he can argue, with Mill, that persons should be free to make their own choices because it makes them happy.[4] The Utilitarian regards agency as an important form of experience; he includes actions and activities among the things that *happen to us*. This is characteristic of the empiricist tradition in which Utilitarianism has its roots, and is nowhere more evident than when Hume writes:

> I desire it may be observ'd, that by the *will*, I mean nothing but *the internal impression we feel and are conscious of, when we knowingly give rise to any new motion of our body, or new perception of our mind*.[5]

Hume here identifies the will not with our power to initiate action, but with the feeling we experience when we exercise that power.

And of course our actions and activities are among the things we experience. But in an equally undeniable sense, having experiences is among the things that we *do*. Activity and passivity are aspects of our nature, not parts, and each can be reduced to a form of the other. In this paper, however, I argue that from a moral point of view it is important not to reduce agency to a mere form of experience. It is important because our conception of what a person is depends on our conception of ourselves as agents in a deep way. My argument is directed against the views about personal identity advanced

by Derek Parfit in *Reasons and Persons*.[6] I believe that Parfit's arguments depend on viewing the person primarily as a locus of experience, and agency as a form of experience. If we regard persons primarily as agents, we will reach different conclusions both about the nature of personal identity and about its moral implications.[7]

II. *Parfit's Theory of Personal Identity*

According to what Parfit calls "the standard view" of identity, the persistence of an object over time can be understood in terms of spatio-temporal continuity under a concept.(203) So long as we can draw a continuous line through space-time, at every point on which there is an object falling under the concept X, we say that the object is the same X, persisting. It is generally acknowledged that this criterion is loose and admits of many exceptions. Some objects may be dismantled and reconstructed later on.(203) Or, an object that does persist through space-time may change so radically that we will not call it the same one even though it falls under the same concept. Finally, the indeterminacy of our concepts often leaves us puzzled whether to say that an object falling under a given concept still exists or not.[8] In such cases, what we say about identity must be governed by pragmatic considerations. There is no metaphysical fact of the matter about whether the earlier object and the later one are the same. The question whether the thing is the same is, as Parfit says, an empty question. (213) It may have a best answer, if the pragmatic considerations are decisive, but it does not have a true answer. (260)

The standard view works best for physical objects. According to the physical criterion of personal identity, a person is, or coincides exactly with, a spatio-temporally continuous living human body. But some people think that physical continuity is neither necessary nor sufficient for personal identity. It is insufficient, because a human being could conceivably change so radically that he would not really be the same person. A great deal of what matters to us in ourselves and others is psychological: our memories,

characters, tastes, interests, loves, hates, and so forth. To take an extreme case, a human being who lost all of his memories, and began to exhibit a very different character, would not be, in any important sense, the same person. For similar reasons, some think that physical continuity is not necessary for identity. They think it is conceivable that the very same psychological person could occupy a different body, or, as in some religious conceptions, that a person could become independent of a body altogether. These people believe that a different kind of continuity characterizes personal identity, which Parfit calls psychological continuity.

Parfit defines psychological continuity in terms of another relation, psychological connection. (205-206) A psychological connection exists when a psychological state at one time is causally related in an appropriate way to a psychological state at an earlier time. For instance, if Marilyn remembers something because Norma Jean experienced it; if Marilyn does something because Norma Jean intended it; if Marilyn believes something because Norma Jean was taught it; then Marilyn and Norma Jean are psychologically connected.[9] If there are many such connections they are strongly connected, and if there are overlapping chains of strong connection then Marilyn is psychologically continuous with Norma Jean. In this case, what Parfit calls "Relation-R" - psychological connectedness and continuity - holds between Marilyn and Norma Jean. Under normal circumstances, this means that Marilyn is Norma Jean at a later date.

There are two views about what makes psychological connectedness possible. Some people believe that psychological connections hold because of something psychological that persists in the person. The easiest way to conceive this is dualistically: the person has a soul or a Cartesian Ego that is the common subject of all of her experiences and links them. Or, Parfit suggests that even without being a dualist, one may believe that the persistence of a psychological subject is a "deep further fact" about a person. Parfit calls these views "Non-Reductionist", since the persisting subject of experiences is regarded as an irreducible entity. On a Non-Reductionist view, the

person is the psychological subject of experiences; the holding of Relation-R results from identity, rather than being constitutive of it.[10] But a physicalist is more likely to believe that psychological connectedness can be explained simply by physical continuity. No persisting psychological *entity* is needed to explain connectedness. Parfit calls this view "Reductionist", since the existence of a person can be reduced to "the existence of a brain and body, and the occurrence of a series of interrelated physical and mental events." (211) When this view is combined with the view that Relation-R is more important to personal identity than physical continuity, Relation-R becomes constitutive of personal identity.[11]

It is important to notice that for the Reductionist who favors a psychological criterion of the person, physical continuity is neither necessary nor sufficient for personal identity.[12] The importance of physical continuity is that it is the *normal* cause of psychological connectedness and so of psychological continuity. But Parfit argues convincingly that nothing important depends on the cause being normal. Even if physicalism is true, the physical basis of our psychological attributes will lie in what we may call, in a broadly Aristotelian sense, the "formal" rather than the "material" properties of the nervous system.[13] That is, it will rest in something about the way the matter is *organized*, not in the particular matter used. If this is correct, psychological connectedness can have other causes. Suppose you are about to lose your memory because a portion of your brain is dying. But a surgeon can make a copy of this portion of your brain, including the configuration that carries memory but without the disease, and substitute it for the dying part. For you, this should be just as good as remembering in the normal way. After all, this is what the rest of your body is always doing: replacing old parts with fresh ones like them. The continuity appropriate to a living being is this "formal" continuity.[14] Indeed, Parfit argues, what we normally count as persisting identity is simply formal continuity plus uniqueness – that is, being the only formal continuer of a past self.

But it is the continuity and not the uniqueness that really matters. For if the essential aspect of a person is formal or copiable, then in principle it seems that we could make two equally good copies of a whole person. Both of them would be formally continuous with the original person; each would have that person's memories, character, ambitions and loves; each would believe, or at least feel, that he is the original person. In this case, the question which of them is the same person as the original is an empty question. And it may not matter, or it may depend on circumstances, what we ought to say about it. And so personal identity is not important. Rather, Relation-R is what matters.

But Relation-R, unlike identity, is a relation we can bear to more than one person. If this is what is important, what matters to me in my survival is not whether "I" survive, but whether someone who is sufficiently R-related to me does. Whether or not that person is *me* may not be important; it may even, in science-fictional cases, be an empty question. But we do not need to appeal to science fiction to make the point. We need only grant to Parfit that it does not matter what the cause of connectedness is. I may do something because I intended it; but so may my lawyer, secretary, executor, or slave. I may believe something because I convinced myself of it; so, indirectly, may my students, children, or friends. Although others will not directly remember events because they happened to me, they may certainly know of events because they happened to me. To the extent that such connectedness, and not identity, is what matters to us in our survival, the second kind of connection may be nearly as good as the first.[15] Yet the second kind of connection is consistent with my death.

Parfit believes that his arguments show that the very existence of "persons" as a distinct kind of entity is less "deep" than we normally take it to be. Reductionists hold that a person just consists in "the existence of a brain and body, and the occurrence of a series of interrelated physical and mental events." (211) They also hold that:

> … these facts can be described without presupposing the identity of this person, or explicitly claiming that the experiences in this person's life are had by this person, or even explicitly claiming that this person exists. These facts can be described in an *impersonal* way. (211)

Parfit believes that the primary challenge to the possibility of an impersonal description is the claim that there must be a *subject* of experiences. To this, he replies:

> In one sense this is clearly true. Even Reductionists do not deny that people exist. And, on our concept of a person, people are not thoughts and acts. They are thinkers and agents. I am not a series of experiences, but the person who *has* these experiences. A Reductionist can admit that, in this sense, a person is *what has* experiences, or the *subject of experiences*. This is true because of the way in which we talk. What a Reductionist denies is that the subject of experiences is a separately existing entity, distinct from a brain and body, and a series of physical and mental events.(223)

Parfit believes that accepting Reductionism will change our minds about many issues concerning rationality and morality. Non-Reductionists think that personal identity constitutes a *special* relation I have to my past and future selves. I am responsible for my past self, and I bear the guilt for her crimes and the obligations created by her promises. I am responsible to my future self, for whose happiness, since it will one day be mine, it is rational for me to provide. Parfit thinks that such beliefs result from thinking of the self as a separately existing entity, a subject of experiences or Cartesian Ego, which moves through the course of a life, collecting memories and responsibilities as it goes. If there is no continuing subject of experiences, he argues, these claims about a special relation to oneself, and their rational and moral implications, are less plausible. Relation-R may bear the weight of some of these special relations, but it is a relation we may have to various persons and to varying degrees. Alternatively,

Relation-R may not establish *any* special relation between myself and the subject of experiences who will occupy my body in the future. In either case, there should be changes in our views about morality and rationality, wherever those depend on claims about personal identity.

III. *The Unity of Agency*

Suppose that Parfit has established that there is no deep sense in which I am identical to the subject of experiences who will occupy my body in the future.[16] In this section I will argue that I nevertheless have reasons for regarding myself as the same rational agent as the one who will occupy my body in the future. These reasons are not metaphysical, but practical.

To see this, first set aside the problem of identity over time, and think about the problem of identity at any given time. Why do you think of yourself as one person now? This problem should seem especially pressing if Parfit has convinced you that you are not unified by a Cartesian Ego which provides a common subject for all of your experiences. Just now you are reading this paper. You may also be sitting in a chair, tapping your foot, and feeling hot or tired or thirsty. But what makes it one person, who is doing and experiencing all this? We can add to this a set of characteristics which you attribute to yourself, but which have only an indirect bearing on your conscious experiences at any given time. You have loves, interests, ambitions, virtues, vices and plans. You are a conglomerate of parts, dispositions, activities, and experiences. As Hume says, you are a bundle.[17] What make you one person even at one time?

In *On the Soul*, Aristotle says that the practical faculty of the soul must be one thing.[18] We think of it as having parts, of course, because we sometimes have appetites that are contrary to practical reason, or experience conflict among our various desires. Still, the faculty that originates motion must be regarded as a single thing, because we do

act. Somehow, the conflicts are resolved, and no matter how many different things you want to do, you in fact do one rather than another.

Your conception of yourself as a unified agent is not based on a metaphysical theory, nor on a unity of which you are conscious. Its grounds are practical, and it has two elements. First, there is the raw necessity of eliminating conflict among your various motives. In making his argument for Reductionism, Parfit appeals to a real-life example which has fascinated contemporary philosophers: persons with split brains.(245-246) When the corpus callosum, the network of nerves between the two hemispheres of the brain, is cut, the two hemispheres can function separately.[19] In certain experimental situations, they do not work together and appear to be wholly unconscious of each other's activities. These cases suggest that the two hemispheres of the brain are not related in any metaphysically deeper way than, say, two people who are married. They share the same quarters and, with luck, they communicate. Even their characteristic division of labor turns out to be largely conventional, and both can perform most functions. So imagine that the right and left halves of your brain disagree about what to do. Suppose that they do not try to resolve their differences, but each merely sends motor orders, by way of the nervous system, to your limbs. Since the orders are contradictory, the two halves of your body try to do different things.[20] Unless they can come to an agreement, both hemispheres of your brain are ineffectual. Like parties in Rawls's original position, they must come to a unanimous decision somehow. You are a unified person at any given time because you must act, and you have only one body with which to act.

The second element in this pragmatic unity is the unity implicit in the *standpoint* from which you deliberate and choose. It may be that what actually happens when you make a choice is that the strongest of your conflicting desires wins. But that isn't the way you think of it when you deliberate. When you deliberate, it is as if there were something over and above all of your desires, something that is *you*, and that *chooses*

which one to act on. The idea that you choose among your conflicting desires, rather than just waiting to see which one wins, suggests that you have reasons for or against acting on them.[21] And it is these reasons, rather than the desires themselves, which are expressive of your will. The strength of a desire may be counted *by you* as a reason for acting on it; but this is different from *its* simply winning. This means that there is some principle or way of choosing that you regard as being expressive of *yourself*, and which provides reasons that regulate your choices among your desires. To identify with such a principle or way of choosing is to be "a law to yourself," and to be unified as such. This does not require that your agency be located in a separately existing entity or involve a deep metaphysical fact. Instead, it is a practical necessity imposed upon you by the nature of the deliberative standpoint.[22]

It is of course important to notice that the particular way you choose which desires to act on *may* be guided by your beliefs about certain metaphysical facts. Parfit evidently thinks that it should. When he argues about the rationality of future concern, Parfit assumes that my attitude about the desires of the future inhabitant of my body should be based on the metaphysics of personal identity. That is, I should treat a future person's desires as *mine* and so as normative for me if I have some metaphysical reason for supposing that she is *me*.[23] But this argument from the metaphysical facts to normative reasons involves a move from "is" to "ought" which requires justification. I will be arguing shortly that there may be other, more distinctively normative grounds for determining which of my motives are "my own"; metaphysical facts are not the only possible ground for this decision. For now, the important points are these: first, the *need* for identification with some unifying principle or way of choosing is imposed on us by the necessity of making deliberative choices, not by the metaphysical facts. Second, the metaphysical facts do not obviously settle the question: I must still decide whether the consideration that some future person is "me" has some special normative force for

me. It is practical reason that requires me to construct an identity for myself; whether metaphysics is to guide me in this or not is an open question.

The considerations I've adduced so far apply to unification at any given moment, or in the context of any given decision. Now let's see if we can extend them to unity over time. We might start by pointing out that the body which makes you one agent now persists over time, but that is insufficient by itself. The body could still be a series of agents, each unified pragmatically at any given moment. More telling considerations come from the character of the things that human agents actually choose. First of all, as Parfit's critics often point out, most of the things we do that matter to us take up time. Some of the things we do are intelligible only in the context of projects that extend over long periods. This is especially true of the pursuit of our ultimate ends. In choosing our careers, and pursuing our friendships and family lives, we both presuppose and construct a continuity of identity and of agency.[24] On a more mundane level, the habitual actions we perform for the sake of our health presuppose ongoing identity. It is also true that we think of our activities and pursuits as interconnected in various ways; we think that we are carrying out plans of life. In order to carry out a rational plan of life, you need to be one continuing person. You normally think you lead one continuing life because you are one person, but according to this argument the truth is the reverse. You are one continuing person because you have one life to lead.

You may think of it this way: suppose that a succession of rational agents *do* occupy my body. I, the one who exists now, need the cooperation of the others, and they need mine, if together we are going to have any kind of a *life*. The unity of our life is forced upon us, although not deeply, by our shared embodiment, together with our desire to carry on long-term plans and relationships. But actually this is somewhat misleading. To ask why the present self should cooperate with the future ones is to assume that the present self has reasons with which it already identifies, and which are independent of those of later selves. Perhaps it is natural to think of the present self as

necessarily concerned with present satisfaction. But it is mistaken. Your present self must, in order to make deliberative choices, identify with something from which you will derive your reasons, but not necessarily with something present. The sort of thing you identify yourself with may carry you automatically into the future: and I have been suggesting that this will very likely be the case. Indeed, the choice of any action, no matter how trivial, takes you some way into the future. And to the extent that you regulate your choices by identifying yourself as the one who is implementing something like a particular plan of life, you need to identify with your future in order to be *what you are even now*.[25] When the person is viewed as an agent, no clear content can be given to the idea of a merely present self.[26]

Still, Parfit might reply that all this concedes his point about the insignificance of personal identity. The idea that persons are unified as agents shares with Reductionism the implication that personal identity is not very deep. If personal identity is just a prerequisite for coordinating action and carrying out plans, individual human beings do not have to be its possessors. We could, for instance, always act in groups. The answer to this is surely that for many purposes we do; there *are* agents of different sizes in the world. Whenever some group wants or needs to act as a unit, it must form itself into a sort of person: a legal person, say, or a corporation. Parfit himself likes to compare the unity of persons to the unity of nations. A nation, like a person, exists, but it does not amount to anything more than "the existence of its citizens, living together in certain ways, on its territory." (211–212) In a similar way, he suggests, a person just amounts to "the existence of a brain and body, and the occurrence of a series of interrelated physical and mental events." (211) On the view I am advancing, a better comparison would be the state. I am using "nation" here, as Parfit does, for an historical or ethnic entity, naturalistically defined by shared history and traditions; a state, by contrast, is a moral or formal entity, defined by its constitution and deliberative procedures. A state is not merely a group of citizens living on a shared territory. We have a state only where these

citizens have constituted themselves into a single agent. They have, that is, adopted a way of resolving conflicts, making decisions, interacting with other states, and planning together for an ongoing future. For a group of citizens to view themselves as a state, or for us to view them as one, we do not need to posit the state as a separately existing entity. All we need is to grant an authoritative status to certain choices and decisions made by certain citizens or bodies, as its legislative voice. Obviously, a state is not a deep metaphysical entity underlying a nation, but rather something a nation can make of itself. Yet the identity of states, for practical reasons, must be regarded and treated as more determinate than the identity of nations.

But the pragmatic character of the reasons for agent unification does not show that the resulting agencies are not *really* necessary. Pragmatic necessity can be overwhelming. When a group of human beings occupy the same territory, for instance, we have an imperative need to form a unified state. And when a group of psychological functions occupy the same human body, they have an even more imperative need to become a unified person. This is why the human body must be conceived as a unified agent. As things stand, it is the basic kind of agent.

Of course if the technology were different, individual human bodies might not be the basic kind of agent. My argument supports a physical criterion of identity, but only a conditional one. *Given the technology we have now*, the unit of action is a human body. But consider Nagel's concept of a "series-person." Nagel imagines a society in which persons are replicated in new matter once every year after they reach the age of thirty. This prevents them from aging, and barring accidents and incurable diseases, may even make them immortal. (289-290) On my concept, a series-person, who would be able to carry out unified plans and projects, and have ongoing relations with other persons, would be a person.[27] But the fact that the basic unit of action might be different if technology were different is neither here nor there. The relevant necessity is the necessity of acting and living, and it is untouched by mere technological possibilities.

The main point of the argument is this: a focus on agency makes more sense of the notion of personal identity than a focus on experience. There is a necessary connection between agency and unity which requires no metaphysical support.

IV. *The Unity of Consciousness*

Many people will feel that my defense of personal unity simply bypasses what is most unsettling in Parfit's arguments. Parfit's arguments depend on what we may broadly call an "Aristotelian" rather than a "Cartesian" metaphysics of the person. That is, matter is essentially particular; form is essentially copiable; and form is what makes the person what she is, and so is what is important about her. The "Cartesian" metaphysics, by contrast, holds that the important element of a person is something essentially particular and uncopiable, like a Cartesian Ego. What tempts people to believe this is an entrenched intuition that something like a Cartesian Ego serves as the locus of the particular consciousness that is mine and no one else's. And my argument about the unity of agency in no way responds to this intuition.

Parfit writes:

> When I believed that my existence was a further fact, I seemed imprisoned in myself. My life seemed like a glass tunnel, through which I was moving faster every year, and at the end of which there was darkness. When I changed my view, the walls of my glass tunnel disappeared. I now live in the open air.(281)

Parfit's glass tunnel is a good image of the way people think of the unity of consciousness. The sphere of consciousness presents itself as something like a room, a place, a lit-up area, within which we do our thinking, imagining, remembering, and planning, and from out of which we observe the world, the passing scene. It is envisioned as a tunnel or a stream, because we think that one moment of consciousness is somehow directly continuous with others, even when interrupted by deep sleep or

anesthesia. We are inclined to think that memory is a deeper thing than it is, that it is *direct* access to an earlier stage of a continuing self, and not merely one way of knowing what happened. And so we may think of amnesia, not merely as the loss of knowledge, but as a door that blocks an existing place.

The sense that consciousness is in these ways unified supports the idea that consciousness requires a persisting psychological subject. The unity of consciousness is supposed to be explained by attributing all of one's experiences to a single psychological entity. Of course, we may argue that the hypothesis of a unified psychological subject does nothing to *explain* the unity of consciousness. It is simply a figure for or restatement of that unity. Yet the idea of such a subject seems to have explanatory force. It is to challenge this intuition that Parfit brings up the facts about persons with divided brains. People are often upset by these facts because they think that they cannot imagine what it is like to be such a person. When the hemispheres function separately, the person seems to have two streams of consciousness. If consciousness is envisioned as a sort of place, then this is a person who seems to be in two places at the same time. If consciousness requires a subject, then this person's body seems, mysteriously, to have become occupied by two subjects. Here, the hypothesis of a psychological subject brings confusion rather than clarity.

Parfit own suggestion is that the unity of consciousness "… does not need a deep explanation. It is simply a fact that several experiences can be co-conscious, or be the objects of a single state of awareness."(250) Split-brain people simply have experiences which are not co-conscious, and nothing more needs to be said. This seems to me close to the truth but not quite right. Privileging the language of "having experiences" and "states of awareness" gives the misleading impression that we can count the experiences we are now having, or the number of objects of which we are aware, and then ask what unifies them. The language of activities and dispositions enables us to characterize both consciousness and its unity more accurately.[28]

Consciousness, then, is a feature of certain activities which percipient animals can perform. These activities include perceiving; various forms of attending such as looking, listening, and noticing; more intellectual activities like thinking, reflecting, recalling, remembering, and reading; and moving voluntarily. Consciousness is not a state that makes these activities possible, or a qualification of the subject who can do them. It is a feature of *the activities themselves*. It is misleading to say that you must be conscious in order to perform them, because your being able to perform them is all that your being conscious amounts to.

Voluntary motion is an important example because of a distinction that is especially clear in its case. When we move voluntarily, we move consciously. But this is not to say we are conscious that we are moving. Much of the time when we move nothing is further from our minds than *the fact* that we are moving. But of course this does not mean that we move unconsciously, like sleepwalkers. It is crucial, in thinking about these matters, not to confuse *being engaged in a conscious activity* with *being conscious of an activity*. Perhaps such a confusion rests behind Descartes's bizarre idea that non-human animals are unconscious. In the direct, practical sense, an adult hunting animal which is, say, stalking her prey, knows exactly what she is doing. But it would be odd to say that she is aware *of* what she is doing or that she knows anything *about* it. What she is aware of is her environment, the smell of her prey, the grass bending quietly under her feet. The consciousness that is inherent in psychic activities should not be understood as an inner *observing* of those activities, a theoretic state. An animal's consciousness can be entirely practical.

The unity of consciousness consists in one's ability to coordinate and integrate conscious activities. People with split brains cannot integrate these activities in the same way they could before. This would be disconcerting, because the integration itself is not something that we are ordinarily aware of. But it would not make you feel like two people. In fact, such persons learn new ways to integrate their psychic functions, and

appear normal and normally unified in everyday life. It is only in experimental situations that the possibility of unintegrated functioning is even brought to light.[29]

What makes it possible to integrate psychic functions? If this is a causal question, it is a question for neurologists rather than philosophers. But perhaps some will still think there is a conceptual necessity here: that such integration requires a common psychological subject. But think again of persons with split brains. Presumably, in ordinary persons the corpus callosum provides means of communication between the two hemispheres; it transmits signals. When split-brain persons are not in experimental situations, and they function normally, the reason appears to be simply that the two hemispheres are able to communicate by other means than the corpus callosum. For example, if the left hemisphere turns the neck to look at something, the right hemisphere necessarily feels the tug and looks too.[30] Activities, then, may be coordinated when some form of communication takes place between the performers of those activities. But communication certainly does not require a common psychological subject. After all, when they can communicate, two different people can integrate their functions, and, for purposes of a given activity, become a single agent.

Communication and functional integration do not require a common subject of conscious experiences. What they do require, however, is the unity of agency. Again, there are two aspects of this unity. First, there is the raw practical necessity. Sharing a common body, the two hemispheres of my brain, or my various psychic functions, must work together. The "phenomenon" of the unity of consciousness is nothing more than the *lack* of any perceived difficulty in the coordination of psychic functions. To be sure, when I engage in psychic activities *deliberately*, I regard myself as the subject of these activities. *I* think, *I* look, *I* try to remember. But this is just the second element of the unity of agency, the unity inherent in the deliberative standpoint. I regard myself as the employer of my psychic capacities in much the same way that I regard myself as the arbiter among my conflicting desires.

If these reflections are correct, then the unity of consciousness is simply another instance of the unity of agency, which is forced upon us by our embodied nature.

V. *Agency and Identity*

At this point it will be useful to say something about why I take the view which I am advancing to be a Kantian one. Kant believed that as rational beings we may view ourselves from two different standpoints.[31] We may regard ourselves as objects of theoretical understanding, natural phenomena whose behavior may be causally explained and predicted like any other. Or we may regard ourselves as agents, as the thinkers of our thoughts, and the originators of our actions. These two standpoints cannot be completely assimilated to each other, and the way we view ourselves when we occupy one can appear incongruous with the way we view ourselves when we occupy the other. As objects of theoretical study, we see ourselves as wholly determined by natural forces, the mere undergoers of our experiences. Yet as agents, we view ourselves as free and responsible, as the authors of our actions, and the *leaders* of our lives. This incongruity need not become contradiction, so long as we keep in mind that the two views of ourselves spring from two different relations in which we stand to our actions. When we look at our actions from the theoretical standpoint our concern is with their explanation and prediction. When we view them from the practical standpoint our concern is with their justification and choice. These two relations to our actions are equally legitimate, inescapable, and governed by reason, but they are separate. Kant does not assert that it is a matter of theoretical fact that we are agents, that we are free, and that we are responsible. Rather, we must view ourselves in these ways when we occupy the standpoint of practical reason – that is, when we are deciding what to do. This follows from the fact that we must regard ourselves as the causes – the first causes – of the things that we will. And this fundamental attitude is forced upon us by the necessity of making choices, regardless of the theoretical or metaphysical facts.[32]

From the theoretical standpoint, an action may be viewed as just another experience, and the assertion that it has a subject may be, as Parfit says, "because of the way we talk." But from the practical point of view, actions and choices must be viewed as having agents and choosers. This is what *makes* them, in our eyes, our own actions and choices rather than events that befall us. In fact, it is only from the practical point of view that actions and choices can be distinguished from mere "behavior" determined by biological and psychological laws. This does not mean that our existence as agents is asserted as a further fact, or requires a separately existing entity that should be discernable from the theoretical point of view.[33] It is rather that from the practical point of view our relationship to our actions and choices is essentially *authorial*: from it, we view them as *our own*. I believe that when we think about the way in which our own lives matter to us personally, we think of ourselves in this way. We think of living our lives, and even of having our experiences, as something that we *do*. And it is this important feature of our sense of our identity that Parfit's account leaves out.[34]

What sort of difference does this make? To put it in Parfit's terms, it privileges certain kinds of psychological connection over others: roughly speaking, authorial ones. In discussing the events that according to Reductionism comprise a person's life, Parfit introduces the idea of a *boring* event: for instance, the continued existence of a belief or a desire.(211) His point in including these, of course, is to cover the fact that one of the things that makes you the same person at $Time_2$ that you were at $Time_1$ is that certain things about you have remained the same. But we can distinguish beliefs and desires that continue merely because, having been acquired in childhood, they remain unexamined, from beliefs and desires that continue because you have arrived at, been convinced of, decided on, or endorsed them. In an account of personal identity which emphasizes agency or authorship, the latter kind of connection will be regarded as much less boring than the former. This is because beliefs and desires you have actively

arrived at are more truly your own than those which have simply arisen in you (or happen to adhere in a metaphysical entity that is you).[35] Recall Mill's complaint:

> Not only in what concerns others, but in what only concerns themselves, the individual or the family do not ask themselves, what do I prefer? or, what would suit my character and disposition? or, what would allow the best and highest in me to have fair play and enable it to grow and thrive? … I do not mean that they choose what is customary in preference to what suits their own inclination. It does not occur to them to have any inclination except for what is customary. Thus the mind itself is bowed to the yoke: even in what people do for pleasure, conformity is the first thing thought of; they like in crowds…, and are generally without either opinions or feelings of home growth, or *properly their own* … [36]

It is, I think, significant that writers on personal identity often tell stories about mad surgeons who make changes in our memories or characters.[37] These writers usually emphasize the fact that after the surgical intervention we are altered, we have changed. But surely part of what creates the sense of lost identity is that the person is changed by *intervention*, from outside. The stories might affect us differently if we imagined the changes initiated by the person herself, as a result of her own choice. You are not a different person *just* because you are very different.[38] Authorial psychological connectedness is consistent with drastic changes, provided those changes are the result of actions by the person herself or reactions for which she is responsible.[39]

It is important to see how these claims do and do not violate Parfit's thesis that we should not care what the causal mechanism of connection is. (286) Given a suitable understanding of the idea of a causal mechanism, the Kantian can agree. If I can overcome my cowardice by surgery or medication rather than habituation I might prefer to take this less arduous route. So long as an authentic good will is behind my desire for greater courage, and authentic courage is the result, the mechanism should not matter.

But for the Kantian it does matter who is initiating the use of the mechanism. Where I change myself, the sort of continuity needed for identity may be preserved, even if I become very different. Where I am changed by wholly external forces, it is not. This is because the sort of continuity needed for what matters to me in my own personal identity essentially involves my agency.

VI. *The Moral Differences*

Parfit believes that accepting Reductionism should modify many of our views about rationality and morality. In particular, he believes that Reductionism lends support to Utilitarian attitudes about paternalism and distributive justice. In this section, I show how a more agent-centered conception of personal identity blocks the Utilitarian implications which Parfit anticipates. Yet the agent-centered conception of the person shares with Reductionism the idea that persons are not deeply or metaphysically separated. And some modification of conventional philosophical views, therefore, does emerge.

VI.a. *Future Concern and Paternalism*

Parfit's argument that Reductionism lends support to paternalism has two parts. First, he argues that Reductionism grounds a challenge to a standard view of rationality: that we have reason to be equally concerned about all parts of our own future. What matters is not identity but Relation-R, and part of that relation, connectedness, is a matter of degree. There are two possible conclusions to draw about the rationality of special concern about one's own future. What Parfit calls the "Extreme Claim" is that I have *no* reason to be especially concerned about my own future. (More properly speaking, I have no reason of the form "she's me" to be especially concerned about any particular future person.) The "Moderate Claim" is that my personal concern about any future person who is R-related to me may (rationally) be a matter of degree. (307-308)

The disquieting result, according to Parfit, is that we cannot always criticize great imprudence as irrational. There may be no irrationality in my imposing a disproportionate burden on a person who will be R-related to me in the future for the sake of myself now. Even if we accept the "Moderate Claim," my future self may be too weakly connected to me to require great concern on my part. The result is disquieting because great imprudence "ought to be criticized." (318) However we characterize it, most of us agree that there is something wrong about engaging in activities and relationships that pose a bad risk to one's future self-esteem, health, or welfare. Parfit proposes, therefore, that we should regard such conduct, even where not irrational, as immoral. Imposing the burdens of diminished self-respect, ill-health, or misery on your later self should be regarded as wrong in exactly the same way that imposing these burdens on other persons is.

But this, in turn, may change our view about "paternalistic" intervention. Parfit writes:

> The person we coerce might say: 'I may be acting irrationally. But even if I am, that is my affair. If I am only harming myself, I have the right to act irrationally and you have no right to stop me.' This reply has some force. We do not believe that we have a general right to prevent people from acting irrationally. But we do believe that we have a general right to prevent people from acting wrongly. This claim may not apply to minor wrong-doing. But we believe that it cannot be wrong, and would often be our duty, to prevent others from doing what is seriously wrong. Since we ought to believe that great imprudence is seriously wrong, we ought to believe that we should prevent such imprudence, even if this involves coercion. (321)

There is more than one problem with this proposal. First, Parfit bases his analysis on an account of ordinary morality which I believe is mistaken. Most people *already* believe

that great imprudence is morally wrong. The ruined or wasted life, with health, opportunity, and talent squandered, seems to us not merely stupid but reprehensible. And strictures against a lack of proper self-concern also follow from most ethical theories. The Greeks arguably made a form of self-concern the basis of their ethical theories, and unarguably included self-regarding virtues alongside others without hesitation. Of the eighteenth century moralists perhaps only Hutcheson, who thought all virtue grounded in benevolence, was prepared to argue that self-regarding attributes could only be virtues indirectly.[40] For the Utilitarian, lack of self-concern is the cause of needless pain and grief; for the Kantian, it evinces a lack of respect for the humanity in one's own person[41]; for the religious moralist, it is a failure of responsibility for what has been placed in one's special care.

The reason we do not feel entitled to interfere with imprudence is not, as Parfit claims, based on the difference between irrationality and immorality. He is also mistaken when he says that we believe that we may always interfere with immorality unless it is minor. The difference here is rather that between the realm of public right and the realm of private virtue. We enforce public right even when it is trivial; we cannot interfere with private vice even when it is as major as the concerns of human life can be. We may use coercion to prevent you from parking your car in your neighbor's unused private driveway or running a red light on a deserted road. We may not use coercion to prevent you from breaking your lover's heart or demolishing your spouse's self-esteem. A person has a right to the disposition of his driveway, but no one has a right not to have his heart broken.[42] This is certainly not because the latter is a minor wrong, but because of the moral territory we are in.

I have claimed that what matters personally is, or at least essentially involves, the view of myself as an agent, as one who chooses and lives a particular life. And, as things stand, it is *qua* the occupant of this particular body that I live a life, have ongoing relationships, realize ambitions, and carry out plans. So long as I occupy this body and

live this life, I am this rational agent, the same one. As I argued earlier, it is misleading to ask whether my present self has a reason to be concerned with my future selves. This way of talking presupposes that the present self is necessarily interested in the quality of present experiences, and needs a further reason to care for more than that. But insofar as I constitute myself as an agent living a particular life, I will not in this way oppose my present self to future ones. And so I do have a personal reason, whether or not I also have a moral one, to care for my future.

But the weight of this kind of reason for future concern is going to bear against extensive paternalistic intervention. This is not for the standard Utilitarian reasons Parfit mentions – that people should learn from their own mistakes, and are in the best position to know whether their own actions are bad for them. (321) If it matters to me to live my own life, and that includes making my own choices and arriving at my own beliefs, then obviously I will not want others to intervene paternalistically unless it is necessary to prevent me from killing or crippling myself. I can only live my own life to the extent that I am free of such interference. We should be opposed to paternalism, then, not because self-concern lies outside of morality, but because freedom is a condition of living one's own life, or even, as we say, of being one's own person.

But this is not to say that the considerations against deep personal separateness that Parfit and I both endorse have no consequences for the standard philosophical model of rationality. I have suggested that agents come in different sizes, and that the human body is merely the basic one. If we grant that the unity of agency is a reason for future concern, then we should grant that I also have reasons to care for the future of larger agencies of which I am a part. Just as I have a personal concern for my physical future, I may have a *personal* concern for the future of my family, the organization for which I work, a project in which I have been active, or the state of which I am a citizen. In fact our existing attitudes reflect this. We are glad if another country makes difficult changes to secure equality for an oppressed minority, but proud if our own does so; sorry if

another country has recourse to needless military bluster, but ashamed when ours does. We care not just about the purposes involved here, but about our own involvement in them, even where it is distant. And this kind of personal concern often extends to the future of the agencies of which we are a part. The territory of practical reasons is not split into two domains – self-interested rationality concerned with the occupant of this particular body on the one hand, and reasons of impartial morality on the other. Instead, the personal concern which begins with one's life in a particular body finds its place in ever-widening spheres of agency and enterprise, developing finally into a *personal* concern for the impersonal: a concern, that is to say, for the fate of one's fellow creatures, considered merely as such.

VI.b. *Compensation and Distributive Justice*

Parfit's treatment of distributive justice begins from consideration of an objection to Utilitarianism advanced by Rawls. (329ff.) In burdening one person in order to benefit another, Rawls argues, Utilitarianism improperly treats social choice as if it were just like individual choice.[43] The difficulty can be brought out in terms of *compensation*. If I am burdened today in order to get a benefit tomorrow, *I* am compensated. But if I am burdened so that *you* can get a benefit, no one is compensated. Therefore, while burdening myself for a future benefit is rational, burdening one person to benefit another is not. This is part of the reason for what Parfit calls "the objection to balancing," that is, balancing the gains of one person against the losses of another.(337)

Parfit agrees that one person cannot be compensated by a benefit to another. But he thinks that Reductionism makes this fact less important. One formulation of his argument is revealing:

> Even those who object to balancing think that it can be justified to impose burdens on a child for his own greater benefit later in his life. Their claim is that a person's burden, while it can be morally outweighed by benefits

> to him, cannot ever be outweighed by mere benefits to others. This is held to be so even if the benefits are far greater than the burdens. The claim thus gives to the boundaries between lives -- or to the fact of non-identity -- overwhelming significance. It allows within the same life what, over different lives, it totally forbids. (338-339)

Parfit thinks that Reductionism makes this position less rational. But notice his equation of "the boundaries between lives" with "the fact of non-identity." This is explicit in the next paragraph when he says:

> The fact that we live different lives is the fact that we are not the same person. If the fact of personal identity is less deep, so is the fact of non-identity. (339)

But this conclusion does not follow even from Reductionism. Or, rather, it only follows if we also adopt a peculiarly agentless conception of what it is to live a life. Living a life as Parfit sees it is a matter of having a series of experiences. Since the idea of a continuing subject of experiences, as anything more than a grammatical convenience, has been discredited, Parfit supposes that the unity of a life has been discredited as well. He concludes that distributive policies should focus on the quality of experiences rather than on lives. But when living a life is conceived as something done by an agent we do not get this result. Lives conceived of as led by agents may be completely separate even if the unity of those agents is pragmatic rather than metaphysically deep. And if living a life in this sense is what matters, distribution should be over lives, and the agents who lead them. As things stand, the basic leader of a life is a human being, and this is what makes the human being the unit of distribution.[44] If technology changes this - for instance, if series-persons become possible - then the appropriate unit of distribution may change.

But still, one might envisage some change in our views to result from our coming to believe that the fact that we lead separate lives *can be* less deep. Not just human

beings, but marriages, friendships, institutions, and states all have lives. If we suppose that I participate in various lives, then there may be more scope for compensation than the objection to balancing allows. I *may* sometimes be compensated for a personal burden by a benefit to a larger life in which I participate. This is not an unfamiliar or revisionist idea, but is already realized in the attitudes most people have towards their friends and family. The efforts we make for the sake of those we love, and for the sake of keeping our relationships alive, are not regarded as uncompensated burdens, any more than the sacrifices we make today in order to benefit ourselves tomorrow are. I do things for my friend not because I calculate that she will do as much for me, but because she is my friend. This is just as comprehensible a reason, all by itself, as doing something for myself because I am myself. So perhaps it would be all right to impose a burden on me in order to benefit one of my close friends.[45]

But notice that *nothing* in this line of reasoning suggests that I can be compensated for a burden by a benefit to a person whose life is unconnected to, and not part of, my own. Even if personal identity is less deep, and our lives can be connected to those of others in much the same way they are connected within, it does not follow that our lives are equally connected to any lives whatever. And they are not. So a Utilitarian criterion for distribution does not follow from this line of thought.

Still, one might think that even a limited expansion of the scope of possible compensation will change our views about distributive justice. But the argument against paternalism just given bears against this conclusion. And indeed to draw it would be to miss an important part of Rawls's point about the essential difference between private and political decision. Even in the most straightforward case of compensation, where a burden is imposed on a person from which she herself will later benefit, compensation by itself does not do the justificatory work. That I will be compensated may give me a reason to accept a burden; it does not give you a reason to impose one on me. The only reason you have to impose one on me is that *I do accept it*.[46] This fact may be obscured

if we start, as Parfit does, from the example of a child, who is a legitimate object of paternalism, and on whose acceptance we cannot wait. In the case of an adult, it is the acceptance, not the compensation, that does the justificatory work.

VII. *Conclusion*

Some of the discussion of Parfit's work has revolved around the question whether we can, or even should, use a morally neutral, metaphysical conception of the person to support one moral theory over others.[47] I believe that the answer depends on what "morally neutral" is taken to mean. When we say a conception is morally neutral, we may mean that it is constructed without regard to the fact that we are going to employ it in moral thinking; or, we may mean that it is constructed without prior dependence on any particular moral theory. I see no point in being neutral with respect to the purposes of moral thinking, nor do I see that metaphysics achieves that kind of neutrality any better than, say, psychoanalysis or biology.[48] On the other hand, if we are to find a basis for deciding among competing moral theories, an initial neutrality with respect to particular theories might be worth having. But Parfit's conception of the person does not have this kind of moral neutrality.

According to Parfit, Utilitarians disagree with those who insist on compensation and other distributive values because Utilitarians think that the question "to whom does it happen?" is like the question "when does it happen?" They regard both of these as "mere differences in position." (340) Reductionism supports this parallel between the two questions because the Reductionist holds that an impersonal description of life is possible. Persons can be said to exist, but, according to Parfit, "this is true only because we describe our lives by ascribing thoughts and actions to people." (341) It is a matter of grammatical convenience. Therefore:

> 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. (341)

So Parfit thinks that Reductionism supports the thesis that the quality of experiences is what matters, and so supports a Utilitarian theory of value. But I believe instead that Parfit has assumed this theory of value from the start. The metaphysical argument about whether a person is a separately existing subject of experiences, or merely a stream of experiences with no separately existing subject, is preceded by an essentially *moral* assumption: the assumption that life is a series of experiences, and so that a person is first and foremost a locus of experiences. If you begin with the view that a person is a subject of experiences, and take away the subject, you are indeed left with nothing but experiences. But you will only begin with that view if you assume from the start that having experiences is what life is all about.

This assumption dictates the reduction of agency to a mere form of experience which I described at the beginning of this paper. That is, it involves regarding our actions and activities as among the things that happen to us, and so, once the subject is removed, as simply among the things that happen. Because they regard doings as mere happenings, Parfit and other Utilitarians suppose that the question "who does it?" is like the question "to whom does it happen?": according to them, merely about position.[49] But from the deliberative standpoint our relationship to our actions and our lives is not merely one of position. It is essential to us that our actions are our own, and we regard living our lives as something that we do.

Unless persons are separately existing entities, Parfit supposes, the ascription of actions to people is a matter of mere grammatical convenience. The Kantian reply is that neither metaphysics nor grammar is the basis for such ascriptions. Rather, the conception of ourselves as agents is fundamental to the standpoint of practical reason, the standpoint from choices are made. And it is from this standpoint that we ask moral

questions, and seek help from moral philosophy. This makes the conception of the agent, along with its unity, an appropriate one to employ in moral thinking. In fact, it is from the standpoint of practical reason that moral thought and moral concepts – including the concept of the person – are generated.

---

[1] I am pleased to acknowledge my debt to Thomas Hill, Jr. The original stimulus for writing this paper came from an opportunity to comment on his paper "Justifying to Oneself" at the Twenty-Seventh Annual Oberlin Philosophy Colloquium in April, 1987, and his paper suggested some of the ideas in it. Section III contains a version of some of my comments on that occasion. I have also had the benefit of extensive comments on earlier versions of the paper from Peter Hylton, Steven Wagner, John Broome, and the referees for *Philosophy and Public Affairs*; colloquium discussions with the philosophy departments at Princeton University and Union College, and discussions on the material in Section IV with Jay Schleusener. I would like to thank them all.

[2] I borrow the term from Samuel Scheffler's *The Rejection of Consequentialism* (Oxford: Clarendon Press, 1982). Thomas Nagel, who also discusses such restrictions in *The View from Nowhere* (New York: Oxford University Press, 1986), calls them "deontological." But that term may be misleading: the mark of such restrictions is not merely that they are not aimed at good consequences. Even Utilitarians accept non-consequentialist requirements as indirect ways to maximize utility. And other consequentialists may count the performance of a required action as being itself a good consequence: when you tell the truth there is a good outcome, an act of honesty has occurred. What is distinctive of agent-centered restrictions is that you are not supposed to violate them even in order to prevent other similar violations. Say, you are to tell the truth even if doing so will have the foreseeable result that several other people will tell lies, while your lying would prevent them from lying. This shows that neither the indirect utility of truth-telling nor the occurrence of acts of honesty exhausts the value of truth-telling for you; you must be in a special way concerned with what *you* do. This may seem paradoxical or self-indulgent to consequentialists. Scheffler, for instance, describes it as *prima facie* irrational.(*loc. cit.* 83) Kantians, however, reject the consequentialist description of the

situation as morally misleading. It is inconsistent with respect for others to regard their actions as simply a *consequence* of what *you* do.

[3] John Rawls, "Justice as Reciprocity" in Samuel Gorovitz, ed., *Utilitarianism with Critical Essays*. (Indianapolis: Bobbs-Merrill, 1971): 242-268, quotation from 267.

[4] See especially *On Liberty*, Chapter III.

[5] David Hume, *A Treatise of Human Nature*, ed. L.A. Selby-Bigge and P.H. Nidditch. (Oxford: Clarendon Press, 1978): 399.

[6] Oxford: Clarendon Press, 1984. Page references to this book are given in parentheses in the text.

[7] Several commentators have noticed that Parfit's conclusions do not square well with the conception of persons as agents. See for instance Samuel Scheffler, "Ethics, Personal Identity, and Ideals of the Person" and Stephen Darwall's reply, "Scheffler on Morality and Ideals of the Person" in *Canadian Journal of Philosophy* 12 (June 1982): 229-264. See also Bart Gruzalski "Parfit's Impact on Utilitarianism" *Ethics* 96 (July 1986): 760-783.

[8] While I draw on Parfit's account of the standard view, mine is a little different. In particular, the idea that identity is continuity under a concept, and that some spurious identity problems arise from the indeterminacy of concepts, is not emphasized in his account. Yet it does come up when he discusses the possibility that his own views have implications for the morality of abortion and euthanasia. (321-323) In these cases, the empty question is not whether a fetus or a comatose body is or is not some particular person, but whether they are persons at all.

[9] It may be urged that a concept like "remember" conceptually implies underlying identity; in a strict sense, I can only "remember" what happened to me. To cover this conceptual objection, Parfit introduces the idea of a "quasi-memory" (or quasi-intention, or whatever). If someone else's memory traces are copied into my brain, say, so that I

seem to remember things that happened to her, we can say that I quasi-remember those things. (219-222) If Parfit is correct in his claim, discussed shortly, that any cause is appropriate for connection, then quasi-memory will count as a psychological connection of the right sort for psychological continuity.

[10] Some further account must then be given, of course, of the continuing identity of the psychological subject of experiences. One of the standard complaints about dualism might be thought to apply, *mutatis mutandis*, here: to say that I remain the same because of a psychic entity within me that remains the same is like saying I am conscious because of a psychic entity within me that is conscious. It only pushes the problem back.

[11] As Parfit himself is at pains to point out (e.g. 209, 241), it is not logically necessary that a dualist be a Non-Reductionist or that a physicalist be a Reductionist. The dualist may believe that the person still reduces to "a series of interrelated physical and mental events"; the physicalist may hold that the person's existence as a continuing subject of experiences is a "deep further fact", even though that subject is not a separate or special kind of substance. Still, physicalism makes Reductionism a more natural view to hold, and I think that that it is helpful to keep this in mind.

[12] In the text I argue that the same person could occupy a different body. Could the same body come to have a new person in it? Suppose psychological changes are so extensive that we say that the person has become "someone else." The physical changes that necessarily accompany these drastic psychological changes need not themselves be so drastic that we will feel like saying that even the body is a new one. This seems obvious, and it is partly on the basis of this intuition that the psychological criterion of identity is favored. But it is worth remembering that the obviousness of the intuition depends on whether we classify the *expressive* features of the body as physical or psychological. If we classify them as physical, then we should expect the physical changes accompanying drastic psychological changes to be drastic as well. Someone who

has begun to walk, sit, stand, hold her limbs and facial muscles differently might without too much strain be said "even to have a different body." In any event, the phenomenon of expressiveness should remind us that the physical/psychological dichotomy is not as hard and fast as we are inclined to suppose.

[13] The use of the Aristotelian terminology here and throughout this section is mine, not Parfit's.

[14] We do not need to appeal to the importance of psychological continuity to establish this conclusion, since it may be argued that it is only the formal aspect of even our physical attributes that matters to us. For more detailed arguments against the importance of material persistence, see Parfit's Chapter 13 and appendix D, "Nagel's Brain." (468-477)

[15] It is not clear to me whether Parfit would agree with this. He defines Relation-R in terms of *direct* connections, without specifying what he means by "direct." Intuitively, knowing something because you remember it does seem more "direct" than knowing something because you are told. In the second case there is an "extra" step - another person must first remember and then tell you. But Parfit also argues that we should accept any sort of cause of psychological connectedness and continuity. His examples of the other sort of causes up for consideration are things like having someone else's memory traces copied into your brain. It is not clear to me that this *is* a more direct way of knowing something than being told. No doubt it would have a more direct *feel* to it, since it would feel like a memory, but that should make no difference. There is still an extra step: the other person must first remember and then the memory is transplanted into your brain.

[16] This formulation is not, I believe, quite right. Parfit's arguments show that there is not a one-one correspondence between persons and human animals, but of course there is no implication that a person ever exists apart from a human animal. So perhaps we

should say that what his arguments show is that the subject of *present* experiences is not the person, but the animal on whom the person supervenes. There are several difficulties with this way of talking, for there are pressures to attribute experiences to the person, not to the animal. It is the person to whom we attribute memory of the experience, and what the person remembers is "such and such happened to me" not "such and such happened to the animal who I was then." And, to the extent that the character of your experiences is conditioned by your memories and character, we should say that the character of your experiences is more determined by which person you are than by which animal you are.(See note 28.) In fact, however, none of this blocks the conclusion that the animal is the subject of experiences in the sense of being immediately conscious of them when they are present. And in this paper, I will be suggesting that we attribute experiences to the person in a different sense: the person is the agent in whose activities these experiences figure, the one who is engaged in having them. It is only if we insist on saying that the person and not the animal is the conscious subject of present experiences that we can get the conclusion in the text.

[17] *Treatise of Human Nature*, 252. Hume, however, would not accept the description of the problem I have just given, for two reasons. First, he thinks that we do not experience more than one thing at a time, but rather that our perceptions "succeed each other with an inconceivable rapidity." (252) Second, he is only talking about the persistence of a subject of "perceptions", or as he puts it "personal identity, as it regards our thought or imagination" which he separates from personal identity "as it regards our passions or the concern we take in ourselves." (253) Taken together, these two points leave Hume with only the diachronic problem of what links a perception to those that succeed and follow it.

[18] Aristotle, *On the Soul*, III.9 -10

---

[19] In my account of these persons, I rely on Thomas Nagel's "Brain Bisection and the Unity of Consciousness" *Synthese* 20 (1971), rpt. in *Mortal Questions* (Cambridge: Cambridge University Press, 1979): 147–164.

[20] This is not an entirely fantastic idea. In one case, a man with a split brain attempted to push his wife away with one hand while reaching out to embrace her with the other. See Parfit 246; and Nagel, "Brain Bisection and the Unity of Consciousness" *loc. cit.* 154.

[21] See Stephen Darwall, "Unified Agency" Chapter 9 of *Impartial Reason* (Ithaca: Cornell University Press, 1983): 101–113.

[22] The problem of personal identity often gets compared to the problem of free will, as both are metaphysical issues that bear on ethics. I hope it is clear from the above discussion that there is another similarity between them. The conception of myself as one, and the conception of myself as free (at least free to choose among my desires), are both features of the deliberative standpoint. And from this standpoint both conceptions find expression in my identification with some principle or way of choosing.

[23] This view is also found in Sidgwick. When Sidgwick attempts to adjudicate between egoistic and Utilitarian conceptions of practical reason, the consideration that favors egoism is this: "It would be contrary to Common Sense to deny that the distinction between any one individual and any other is real and fundamental, and that consequently 'I' am concerned with the quality of my existence as an individual in a sense, fundamentally important, in which I am not concerned with the quality of the existence of other individuals: and this being so, I do not see how it can be proved that this distinction is not to be taken as fundamental in determining the ultimate end of rational action." (*The Methods of Ethics* (Indianapolis: Hackett, 1981): 498). But the Utilitarian, appealing to metaphysics rather than common sense, replies: "Grant that the Ego is merely a system of coherent phenomena, that the permanent identical 'I' is not a fact but a fiction, as Hume and his followers maintain; why, then, should one part of the

series of feelings into which the Ego is resolved be concerned with another part of the same series, any more than with any other series?" (*loc. cit.* 419) Parfit endorses the basic form of Sidgwick's argument explicitly on 139. Neither Sidgwick nor Parfit shows why these metaphysical views are supposed to have the normative force suggested.

[24] As Susan Wolf points out "Love and moral character require more than a few minutes. More to the point, love and moral character as they occur in the actual world occur in persons, or at any rate in psychophysical entities of some substantial duration…" "Self-Interest and Interest in Selves" *Ethics* 96 (July 1986): 704-720; quotation from 709.

[25] This way of looking at things places a constraint on how we formulate the reasons we have for desiring to carry on long-term projects and relationships. We cannot say that we want them because we expect to survive for a long time; instead, these things give us reasons for surviving. So the reasons for them must be independent of expected survival. See Bernard Williams, "Persons, Character, and Morality" in *Moral Luck* (Cambridge: Cambridge University Press, 1981: 1-19), especially the discussion of Parfit at 8-12.

[26] I would like to thank a referee for *Philosophy and Public Affairs* for prompting me to be clearer on this point.

[27] On the other hand, Williams's person-types (Parfit 293-297), of whom a number of copies (tokens) exist simultaneously, are not persons, since the tokens would not necessarily lead a common life. See Parfit 293-297; Bernard Williams "Are Persons Bodies?" in *The Philosophy of the Body* , ed. Spicker (Chicago: Quadrant Books, 1970); rpt. in Williams, *Problems of the Self*, (Cambridge: Cambridge University Press: 1973): 64-81.

[28] I have argued that the idea of a momentary agent is unintelligible; I will now suggest, perhaps more surprisingly, that even the idea of a momentary experience is suspect. Consider, for instance, what seems to be one of the clearest cases of a temporally

localized experience: physical pain. There is a clear sense in which pain is worse if you have been in pain for a long while. If pain is a momentary experience, we must suppose that this particular form of badness can be cashed out in terms of the quality of the experience you are having now; so that, I suppose, a clever brain surgeon by stimulating the right set of nerves could make you have exactly the experience of a person who has been in pain for a long while even if you have not. The idea that the intrinsic goodness or badness of an experience can always be cashed out in terms of the felt quality of the experience at the time of having it is explicitly defended in Sidgwick's *Methods of Ethics*, Book II, Chapters II–III, and Book Book III, Chapter XIV. I do not think Sidgwick's arguments are successful, but at least he sees that the point needs defending. A more complex challenge to Sidgwick's thesis comes from the fact that there is a sense in which a pain (I feel like saying: the *same* pain) can be worse if in the face of it you panic, or lose your sense of humor, or give way to it completely. And this will be determined not just by how bad the pain is, but by your character. There is a kind of courage that has to do with how one handles pain, and this suggest that even "experiencing pain" is something that can be *done* in various ways. Privileging the language of conscious states or experiences can cause us to overlook these complications.

[29] Nagel, in "Brain Bisection and the Unity of Consciousness," also arrives at the conclusion that the unity of consciousness is a matter of functional integration, but he believes that there is something unintuitive or unsatisfactory about thinking of ourselves in this way.

[30] Nagel, "Brain Bisection and the Unity of Consciousness" *loc. cit.* 154.

[31] No single reference is adequate, for it is a conception that unfolds throughout Kant's writings. But for the most explicit account of the two standpoints view see the *Foundations of the Metaphysics of Morals*, Part III.

[32] Some people suppose that this means that freedom and agency are an *illusion* produced by the practical standpoint. But this presupposes the primacy of the theoretical standpoint, which is in fact the point at issue. Free agency, and, according to my argument, unified personal identity, are what Kant calls "Postulates of Practical Reason." (See *The Critique of Practical Reason*, trans. Lewis White Beck (Bobbs-Merrill (now Macmillan) Library of Liberal Arts, 1956): 137ff; Prussian Academy pages 132ff.

[33] Contrary to the view of Bart Gruzalski, in "Parfit's Impact on Utilitarianism." Gruzalski claims that a deep further fact is required to support any conception of agency more libertarian than Hume's. (*loc. cit.* 767.)

[34] That it is lives and not merely experiences that matter, and that lives cannot be understood merely as sequences of experiences, is a point that several of Parfit's commentators have made. Thus Wolf urges that "…the value of these experiences depends on their relation to the lives of the persons whose experiences these are." ("Self-Interest and Interest in Selves" *loc. cit.* 709) And Darwall, commenting on Scheffler's response to Parfit, emphasizes the difference between "a conception of the kind of life one would like oneself and others to lead as opposed to the kind of things that befall people." ("Scheffler on Morality and Ideals of the Person" *loc. cit.* 249-250).

[35] Other critics of Parfit have stressed the importance of what I am calling the authorial connection. Darwall, in "Scheffler on Morality and Ideals of the Person," reminds us that "The capacity to choose our ends, and rationally to criticize and assess even many of our desires, means that our future intentions and desires do not simply befall us; rather, they are to some degree in our own hands." (*loc. cit.* 254) And in "Self-Interest and Interest in Selves" Wolf writes, "Being a rational agent involves recognizing one's ability to make one's own decisions, form one's own intentions, and plan for one's own future…" (*loc. cit.* 719) Alternatively, a desire or a belief that has simply arisen in you may be reflectively endorsed, and this makes it, in the present sense, more authentically your

own. See Harry Frankfurt's papers "Freedom of the Will and the Concept of a Person" *Journal of Philosophy* 68 (January 1971): 5-20; "Identification and Externality" in Amélie Rorty (ed.) *The Identities of Persons* (Berkeley: University of California, 1976: 239-251); and "Identification and Wholeheartedness" in Ferdinand Schoeman (ed.) *Responsibility, Character, and the Emotions: New Essays in Moral Psychology* (Cambridge University Press, 1988): 27-45.

[36] John Stuart Mill, *On Liberty*. (Indianapolis: Hackett, 1978): 58-59. My emphasis.

[37] Some of Parfit's own stories involve surgical intervention and in this he follows Williams in "The Self and the Future" (*Philosophical Review* 79 (1970); reprinted in *Problems of the Self* (Cambridge: Cambridge University Press, 1973): 46-63). It is also significant, in a related way, that these writers focus on the question of future physical pains. Although it is true that there is an important way in which my physical pains seem to happen to *me* and no one else, it is also true that they seem to have less to do with who I am (which *person* I am) than almost any other psychic events. (But see note 28 for an important qualification of this remark.) The *impersonal* character of pain is part of what makes it seem so intrusive. Williams uses pain examples to show you how strongly you identify with your body. One might say, more properly, that they show you how strongly you identify with the animal who you (also) are. It is important to remember that we each have an animal identity as well as our more specifically human identity and that some of the most important problems of personal integration come from this fact. (See note 16.) One might say, a little extravagantly, that the growing human animal is disciplined, frustrated, beaten, and shaped until it becomes a person - and then the person is faced with the task of re-integrating the animal and its needs back into a human life. That we are not much good at this is suggested by psychoanalytic theory and the long human history of ambivalence (to say the least) about our bodily nature. Pain examples serve to show us how vulnerable our animal identity can make our human identity.

[38] One of the few things I take issue with in Wolf's "Self-Interest and Interest in Selves" is a suggestion that persons who regarded themselves as R-related to rather than identical with their future selves would be less likely to risk projects that might involve great psychological change. Wolf reasons that great changes would be viewed as akin to death.(*loc. cit.* 712) It should be clear from the above that I think it depends on how one envisages the changes arising.

[39] Parfit does notice the difference between deliberate changes and those brought about by "abnormal interference, such as direct tampering with the brain" (207) but he seems to take it for granted that those who feel that identity is threatened by the latter kind of changes are concerned about the fact that they are *abnormal*, not the fact that they are *interference*.

Of course the sorts of considerations that feed worries about free will and determinism make it hard to distinguish cases in which a person has been changed by external forces from cases in which she has changed herself. Surgical intervention seems like a clear case of external interference because the person's prior character plays no role in producing the result. But what of someone who changes drastically in response to tragedy or trauma? I don't take up these problems here, but only note that from our own perspective we do distinguish cases in which we change our minds, desires, or characters from those in which the changes are imposed from without.

[40] See his *An Inquiry Concerning the Original of Our Ideas of Virtue or Moral Good* (1725), §II. The relevant passages can be found in D.D. Raphael (ed.) *British Moralists 1650-1800* (Oxford: Clarendon Press, 1969), Volume 1: 271.

[41] Of course a Kantian does not believe in the split between rationality and morality that underlies Parfit's analysis in the first place. In a Kantian view, as I have been arguing, no aim is *my own* unless it is the object of my own choice. And if Kant is right in supposing that a choice I may regard as truly my own must also be a universalizable choice, no split

between personal rationality and morality is possible. In other words, Kant supposes that the view of the person which I have been arguing for in this paper leads to the adoption of a particular unifying principle of action, the categorical imperative. I have not tried to argue for this more ambitious thesis here.

[42] I am using "right" here in the strict sense usual in the contract theory tradition, where a right is something that may be coercively enforced: since Parfit is discussing the possible use of coercion, this seems appropriate. Of course, the difference between using coercion and trying to *persuade* the wrongdoer to desist is essential here. But even the latter is normally permitted only to close friends or relatives of the people involved. As Steven Wagner points out to me, we do say things like "You had no right to treat her that way" in private contexts where coercion is not at issue. A Kantian would say that we use this language because of the way the private duties of respect model the public, and enforceable, duties of justice. (When we speak this way, we do not mean merely that it was not *benevolent* to treat her that way.) Although coercion obviously cannot be used to make me respect another, I should regard my respect as something to which she has a claim, just as she does to her rights; respect is not something we give to others out of generosity. See Immanuel Kant, *The Metaphysical Principles of Virtue* in *Kant's Ethical Philosophy*, translated by James Ellington. (Indianapolis: Hackett, 1983): 113-114 and 127-133; Prussian Academy Edition pages (in margins of most translations) 449-450 and 462-468.

[43] John Rawls, *A Theory of Justice* (Cambridge, Massachusetts: Harvard University Press, 1971): §5, especially 26-27.

[44] In actual fact, for economic purposes the unit is often the family, and this is because they are presumed to share their lives, although of course in an economic rather than a metaphysical sense.

[45] Friendship is not a form of altruism. In routine cases, the question of "making a sacrifice" – if that is supposed to be an uncompensated burden – does not even come up. Where a burden is large, one may speak of "making a sacrifice" for one's friend. But then, where a burden is large, one may also speak of "making a sacrifice" for one's career or health. Sometimes, the impossibility of compensation springs from the incommensurability of values, not from who gets what. My friend's happiness may be incommensurable with other things I care about, without being any the less a part of my own happiness for all that.

[46] It will help to recall here that, according to the social contract theory accepted by Kantians, the burdens of social life are supposed to be ones that the citizens accept, through their representatives. Any coercive measure must have this kind of backing.

[47] See for instance: John Rawls, "The Independence of Moral Theory" (*Proceedings and Addresses of the American Philosophical Association* 47 (1974–75): 15–20); Norman Daniels, "Moral Theory and the Plasticity of Persons" (*The Monist* 62 (July 1979): 269); Samuel Scheffler "Ethics, Personal Identity, and Ideals of the Person" (*loc. cit.* 240ff.); and Bart Schultz, "Persons, Selves, and Utilitarianism" *Ethics* 96 (July 1986): 721–745; especially 741ff.

[48] Parfit might reply that the point of appealing to a metaphysical conception is not merely that it is neutral, but that it is deep. It is what we most truly are. But both the truth and the force of this consideration are questionable. Parfit's conception of the person is recognizably metaphysical in that it is concerned with the theoretical conditions of identity and counting, certainly traditional concerns of metaphysics. It is also as minimal as possible. But our metaphysical concerns about countability and ontological economy are still just some concerns among others. And they are not obviously the important ones for ethics.

[49] This is related to the Utilitarian's perplexity about agent-centered restrictions, discussed in note 1. In his discussion of his now well-known example of Jim, who is invited by a South American soldier to kill one Indian in order to save the lives of nineteen others, Bernard Williams says that the Utilitarian solution of the problem regards Jim as "the agent of the satisfaction system who happens to be at *a particular point at a particular time* : in Jim's case, our man in South America." See "A Critique of Utilitarianism," in J.J.C. Smart and Bernard Williams, *Utilitarianism For and Against*, (Cambridge: Cambridge University Press, 1973): 115, my emphasis.