5

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

Christine M. Korsgaard

[...]

II The Unity of Agency

Suppose 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.[1] 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 your experiences. Just now you are reading this article. 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.[2] What makes you one person even at one time?

In *On the Soul,* Aristotle says that the practical faculty of the soul must be one thing.[3] 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 (245–46).[4] When the corpus callosum, the network of nerves between the two hemispheres of the brain, is cut, the two hemispheres can function separately.[5] 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.[6] 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 of 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 is not the way you think of it when you deliberate. When you deliberate, it is as if there were something over and above all 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.[7] 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 expressive of *yourself,* and that 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.[8]

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 concern about the future, 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*.[9] But this argument from the metaphysical facts to normative reasons involves a move from "is" to "ought" which requires justification. I will argue 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 have adduced so far apply to unification at any given moment, or in the context of any given decision. Now let us see whether 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.[10] 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. In order to make deliberative choices, your present self must 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.*[11] When the person is viewed as an agent, no clear content can be given to the idea of a merely present self.[12]

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" (211-12). 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 a 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, they 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 our 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 Thomas 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 (289-90). 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.[15] 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.

III The Unity of Consciousness

Many 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, 1 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 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's 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.[14]

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 perform 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 lies behind Descartes' bizarre idea that nonhuman 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 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.[15]

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.[16] 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 tiy 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.

IV Agency and Identity

At this point it will be useful to say something about why I take the view I am advancing to be a Kantian one. Kant believed that as rational beings we may view ourselves from two different standpoints.[17] 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. The 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.

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 discernible from the theoretical point of view.[19] 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.* 1 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.[20]

What sort of difference does this make? To put it in Parfit's terms, it privileges certain kinds of psychological connection - roughly speaking, authorial ones - over others. 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 inhere in a metaphysical entity that is you).[21] 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.*[22]

It is, I think, significant that writers on personal identity often tell stories about mad surgeons who make changes in our memories or

characters. 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.[24] 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.[25]

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 cormection 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. [...]

Notes

1 This formulation is not, I believe, quite right. Parfit's arguments show that there is not a one-to-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 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 14 below). In fact, however, none of this blocks the conclusion that the animal is the subject of experiences in the sense that it is immediately conscious of them when they are present. And I will suggest 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.

2 [David] Hume, [A] *Treatise of Human Nature,* [ed. L. A. Selby-Bigge and P. H. Nidditch (Oxford: Clarendon Press, 1978),] p. 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" (ibid.). Second, he is talking only 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" (ibid., p. 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.

3 Aristotle, *On the Soul,* III. 9-10.

4 Page numbers in parentheses are to Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984).]

5 In my account of these persons. I rely on Thomas Nagel's "Brain Bisection and the Unity of Consciousness," *Synthese* 20 (1971), repr. in *Moral Questions* (Cambridge: Cambridge University Press, 1979), pp. 147-64.

6 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, *Reasons and Persons,* p. 246, and Nagel, "Brain Bisection and the Unity of Consciousness," in *Moral Questions,* p. 154.

7 See Stephen Darwall, "Unified Agency," in *Impartial Reason* (Ithaca: Cornell University Press, 1983), pp. 101-13.

8 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.

9 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 1 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), p. 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?" (ibid., p. 419). Parfit endorses the basic form of Sidgwick's argument explicitly in *Reasons and Persons,* p. 139. Neither Sidgwick nor Parfit shows why these metaphysical views are supposed to have the normative force suggested.

10 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 (1986): 709).

11 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). pp. 1-19, especially the discussion of Parfit on pp. 8-12.

12 I would like to thank the Editors of *Philosophy & Public Affairs* for prompting me to be clearer on this point.

13 On the other hand. Williams's person-types, of whom a number of copies (tokens) exist simultaneously, are not persons, since the tokens would not necessarily lead a common life. See Parfit, *Reasons and Persons,* pp. 293-7, and Bemard Wilhams, "Are Persons Bodies?," in *The Philosophy of the Body,* ed. Stuart F. Spicker (Chicago: Quadrant Books, 1970). repr. in Bernard Williams. *Problems of the Self* (Cambridge: Cambridge University Press, 1973), pp. 64-81.

14 I have argued that the idea of a momentary agent is unintelligible: I would also like to 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 explicated 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 explicated in terms of the felt quality of the experience at the time of having it is defended in Sidgwick's *Methods of Ethics,* bk. II, chaps. II–III. and bk. III, chap. 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 suggests 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.

15 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.

16 Ibid., in *Mortal Questions,* p. 154.

17 No single reference is adequate, for this conception unfolds throughout Kant's writings. But for the most explicit account of the "two standpoints" view see [Immanuel Kant], *Foundations of the Metaphysics of Morals,* [ed. and trans. R. P. Wolff (Indianapolis: Bobbs-Merill, 1969)] pt. 111.

18 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 (Indianapolis: Bobbs-Merrill, 1956), pp. 137ff; Prussian Academy ed., pp. 132ff).

19 Contrary to the view of Gruzalski in "Parfit's Impact on Utilitarianism," *[Ethics* 96 (1986): 721–45]. Gruzalski claims that a deep further fact is required to support any conception of agency more libertarian than Hume's (ibid., p. 767).

20 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," p. 709). And Darwall, commenting on Scheffler's response to Parfit, emphasizes "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." *[Canadian Journal of Philosophy* 12 (1982): 229-64] pp. 249–50).

21 Other critics of Parfit have stressed the importance of what 1 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" (p. 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" (p. 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, "Freedom of the Will and the Concept of a Person," *Journal of Philosophy* 68 (1971): 5-20; "Identification and Externality," in *The Identities of Persons,* ed. Amélie Rorty (Berkeley and Los Angeles: University of California Press, 1976). pp. 239-51; and "Identification and Wholeheartedness," in *Responsibility, Character, and the Emotions: New Essays in Moral Psychology,* ed. Ferdinand Schoeman (Cambridge: Cambridge University Press, 1988), pp. 27–45. Parfit himself suggests that Reductionism "gives more importance to how we choose to live" *(Reasons and Persons,* p. 446).

22 Mill, *On Liberty* (Indianapolis; Hackett, 1978), pp. 58-9 (emphasis added).

23 Some of Parfit's own stories involve surgical intervention, and in this he follows Bernard Williams in "The Self and the Future," *Philosophical Review* 79 (1970), repr. in *Problems of the Self* pp. 46-63 [and as ch. 1 above]. 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 1 am (which *person* I am) than almost any other psychic events. (But see note 14 above 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 how strongly we identify with our bodies. One might say, more properly, that they show how strongly we identify with the animals who we (also) are. It is important to remember that each of us has 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 1 above). 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 reintegrating 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.

24 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 (ibid., p. 712). It should be clear from the above that I think this depends on how one envisages the changes arising.

Parfit does notice the difference between deliberate changes and those brought about by "abnormal interference, such as direct tampering with the brain" *(Reasons and Persons,* p. 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 do not 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.