Personal Identity

What does being the person that you are, from one day to the next, necessarily consist in? This is the question of personal identity, and it is literally a question of life and death, as the correct answer to it determines which types of changes a person can undergo without ceasing to exist. Personal identity theory is the philosophical confrontation with the most ultimate questions of our own existence: who are we, and is there a life after death? In distinguishing those changes in a person that constitute survival from those changes in a person that constitute death, a criterion of personal identity through time is given. Such a criterion specifies, insofar as that is possible, the necessary and sufficient conditions for the survival of persons.

One popular criterion, associated with Plato, Descartes and a number of world religions, is that persons are immaterial souls or pure egos. On this view, persons have bodies only contingently, not necessarily; so they can live after bodily death. Even though this so-called *Simple View* satisfies certain religious or spiritual predilections, it faces metaphysical and epistemological obstacles, as we shall see.

Another intuitively appealing view, championed by John Locke, holds that personal identity is a matter of *psychological continuity*. According to this view, in order for a person X to survive a particular adventure, it is necessary and sufficient that there exists, at a time after the adventure, a person Y who psychologically evolved out of X. This idea is typically cashed out in terms of overlapping chains of direct psychological connections, as those causal and cognitive connections between beliefs, desires, intentions, experiential memories, character traits, and so forth. This Lockean view is well suited for thought experiments conducted from first-person points of view, such as body swaps or tele-transportation, but it, too, faces obstacles. For example, on this view, it appears to be possible for two future persons to be psychologically continuous with a presently existing person. Can one really become two? In response to this problem, some commentators have suggested that, although our beliefs, memories, and intentions are of utmost importance to us, they are not necessary for our identity, our persistence through time.

A third criterion of personal identity is that we are our bodies, that is to say, that personal identity is constituted by some brute physical relation between, for example, different bodies or different life-sustaining systems at different times. Although this view is still somewhat

unpopular, developments about personal identity theory in the 1990s promise an ideological change, as versions of the so-called somatic criterion, associated with Eric Olson and Paul Snowdon, attract a continuously growing number of adherents.

The aim of this article is to (1) add precision to the problem of personal identity, (2) state a number of theories of personal identity and give arguments for and against them, (3) formulate "the paradox of identity," which proposes to show that posing the persistence question, in conjunction with a number of plausible assumptions, leads to a contradiction, and (4) explain how Derek Parfit's theory of persons attempts to answer this paradox.

Table of Contents

1. Understanding the Problem of Personal Identity

The persistence question, the question of what personal identity over time consists in, is literally a question of life and death: answers to it determine, insofar as that is possible, the conditions under which we survive, or cease to exist in the course of, certain adventures. These adventures do not have to be theoretically as fancy as the cases, to be discussed later, of human fission or brain swaps: a theory of personal identity tells us whether we can live through the acquisition of

complex cognitive capacities in our development from fetus to person, or whether we have survived car accidents if we find ourselves in a persistent vegetative state. Furthermore, theories of personal identity have ethical and metaphysical implications of considerable magnitude: in conjunction with certain normative premises they may support the justification or condemnation of infanticide or euthanasia, or they could prove or falsify certain aspects of our religious outlook, in deciding the questions of how and whether we can be resurrected and whether we are possessors of souls whose existence conditions are identical with ours. It is not surprising, therefore, that most great philosophers have attempted to solve the problem of personal identity, or have committed themselves to metaphysical systems that have substantial implications with regards to the problem, and that most religious belief systems give explicit answers to the persistence question. Neither is it surprising that virtually everybody holds a pre-theoretical theory of personal identity, if only in the sense of having beliefs about afterlives and the meaning of death. The task of solving the metaphysical problem of personal identity essentially involves answering the question of how the phenomenon or principle in virtue of which "entities like us" persist through time is to be specified, under the widely but not universally accepted premises that there is such a phenomenon or principle and that it can be specified. We are concerned, in other words, with the truth-makers of personal identity statements: what makes it true that our statement that an entity X at time t_1 and an entity Y at time t_2 are identical, if X and Y are entities like us?

a. Criteria and the Identity Relation

Answers to the persistence question often provide a criterion of personal identity. A criterion is a set of non-trivial necessary and sufficient conditions that determines, insofar as that is possible, whether distinct temporally indexed person-stages are stages of one and the same continuant person. (A temporally indexed person-stage is a slice of a continuant person that extends in three spatial dimensions but has no temporal extension.) To say that C is a necessary condition for E is to say that if E is the case, then C is the case as well, and to say that C is a sufficient condition for E is to say that if C is the case, then E is the case as well. Consequently, to specify such a criterion is to give an account of what personal identity necessarily consists in.

Let us distinguish between numerical identity and qualitative identity (exact similarity): X and Y are numerically identical iff X and Y are one thing rather than two, while X and Y are

qualitatively identical iff, for the set of non-relational properties $F_1 \ldots F_n$ of X, Y only possesses $F_1 \ldots F_n$. (A property may be called "non-relational" if its being borne by a substance is independent of the relations in which property or substance stand to other properties or substances.) Personal identity is an instance of the relation of numerical identity; investigations into the nature of the former, therefore, must respect the formal properties that govern the latter. The concept of identity is uniquely defined by (a) the logical laws of congruence: if X is identical with Y, then all non-relational properties borne by X are borne by Y, or formally "$\forall(x, y)[(x = y) \rightarrow (Fx = Fy)]$; and (b) reflexivity: every X is identical with itself, or formally "$\forall x(x = x)$. (Note that congruence and reflexivity entail that identity is symmetric, "$\forall(x, y)[(x = y) \rightarrow (y = x)]$, and transitive, "$\forall(x, y, z)[((x = y) \ \& \ (y = z)) \rightarrow (x = z)]$). [Note: $\forall(x, y)$ is an abbreviation of $(\forall x)(\forall y)$.]

Grasp of the notion of numerical identity, to be sure, is essential to our ability to distinguish between the events of picking out one thing more often than once and picking out more than one thing. Although exact similarity is, by congruence, a necessary condition for synchronic personal identity, it is neither necessary nor sufficient for diachronic personal identity, that is to say, the persistence of a person over time: two person-slices at different times could be qualitatively identical slices of different people or qualitatively distinct slices of the same person. This is not to say, however, that it is ruled out that lack of similarity over time may obliterate numerical personal identity: depending on what personal identity consists in, certain qualitative changes in a person's psychology or physiology may kill the person. The question a criterion of personal identity answers is: what kind of changes does a person survive?

This gives a distinctive sense to the claim that a criterion of personal identity is to be constitutive, not merely evidential: in order for a relation *R* to be constitutive for personal identity, it must be the case that, necessarily, if some past or future Y stands in an *R*-relation to X, then X is identical with Y. Hence, many elements of our successful everyday reidentification practices, such as physical appearance, fingerprints, or signatures, are inadequate if considered as constituting ingredients of personal identity relations: for example, if the man in the crowd is wearing a Yankees jacket, this might be sufficient evidence for you to conclude that he is your friend Larry. However, wearing a Yankees jacket is not what it is for Larry to persist through

time: neither did Larry come into existence when he wore the jacket for the first time nor does he die when he takes it off.

Does the logic of the concept of identity impose further restraints on the concept of personal identity? Some commentators believe that identity is an intrinsic relation, that is, that if two person-stages at different times are stages of one and the same person, that will be true only in virtue of the intrinsic relation between these two stages (cf. Noonan 1989; Wiggins 2001). Others hold identity to be necessarily determinate, that is, that it is necessarily false that sometimes there is no answer to the question of whether X is identical with Y. These commentators typically reason as follows: suppose that it is indeterminate that X is identical with Y. Since it is determinate that X is identical with X, under the assumption that congruence and predicate logic apply, X must be determinately identical with Y. Therefore, by *modus tollens*, if X is not determinately identical with Y, X is not identical with Y (cf. Evans 1985; Wiggins 2001). Consequently, the question does in fact have an answer, and the claim that identity is indeterminate is self-contradictory. This conclusion is strengthened, in the case of personal identity, by the widely shared intuition that *even if* the identity of some objects might be indeterminate, this could not be true of the identity of persons: one cannot, it seems, be a bit dead and a bit alive in the same way in which one cannot be a bit pregnant. As it turns out, however, there may be good reasons to deny both the intrinsicness and the determinacy of personal identity (cf. 3.a.; 3.b.).

b. Personhood

While the formal properties of the concept of identity are necessary constraints on our discussion, the truth of our identity judgments is subject to material conditions of correctness, which these formal properties cannot provide. These material conditions must be supplied by the nature of the relata judged to stand in an identity relation. The obvious suggestion is that, given that we are dealing with *personal* identity, these relata are person-stages located at different times. This proposal, however, violates the requirement that the persistence question ought to specify its relata without presupposing an answer: should we choose to accept a definition in the vicinity of Locke's characterization of a person as a "thinking, intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing in different times and places" (1689, II.xxvii.9), then those criteria of personal identity that sanction the identity of a person at one time with a non-person at another time are categorically ruled out. Fetuses, infants,

or human beings in a persistent vegetative state, for example, plainly do not fulfill the criteria envisaged by Locke. As a result, since these beings do not possess cognitive capacities, if they do at all, that qualitatively attain those of thinking beings, couching the persistence question in terms of persons entails that none of us has ever been a fetus or infant or ever will be a human vegetable (Olson 1997a; Mackie 1999). To be sure, these initially baffling claims *could* be true. However, since these are clearly substantial questions about our persistence, we should not consider ourselves justified to settle the matter by definition. Consequently, we should prefer vagueness over chauvinism and pose the persistence question in terms of the wider notion of human being, postponing the question of whether and in what sense the notions of person and human being ought to be distinguished: for any person X and any human being Y at different times t_1 and t_2, if X at t_1 is numerically identical with Y at t_2, what makes this claim necessarily true?

2. Theories of Personal Identity

In order to discover what your pre-philosophical attitude towards this question is, ask yourself the following: what does a supernatural being have to do in order to resurrect you after you die? Collect a few possible answers and ask yourself whether the resulting being, the freshly created being that is now a candidate for being identical with you before you died, is in fact you. For example, do you believe that

1. …the supernatural being could have given you a body which bears no physical continuity or causal relation to the one you possessed before your death, or that it could have resurrected you, in some sense or other, as a bodiless being?
2. …it could have given a new form or content to your psychology, that is, that it is not necessary or sufficient for the "resurrected you" to remember your actions or experiences and that there do not have to be any causal connections between the actions and experiences of you before you died and the"resurrected you"?
3. …the question of whether or not the resulting person is you depends on the existence, in the resurrected person, of something that one might call "a soul"?

If you believe any of these options, then you must also believe, respectively, that

1. …a physiological criterion of personal identity is false.

2. …a psychological criterion of personal identity is false.
3. …the Simple View of personal identity is true.

Let us discuss these theories of personal identity in more detail.

a. The Simple View

Some commentators believe that there are no informative, non-trivial persistence conditions for people, that is, that personal persistence is an ultimate and unanalyzable fact (cf. Chisholm 1976; Lowe 1996; Merricks 1998; Shoemaker & Swinburne 1984). While psychological and physiological continuities are evidential criteria, these do not constitute necessary and/or sufficient conditions for personal identity. We must distinguish between two versions of this view. One version is that personal identity is non-reductive and wholly non-informative, denying that personal identity follows from anything other than itself. This makes the label Identity Mysticism ("*IM*") most appropriate (cf. Zimmerman 1998):

IM: X at t_1 is identical to Y at t_2 iff X at t_1 is identical to Y at t_2,

Identity Mysticism plays only an indirect role in contemporary personal identity theory. Although it may be poorly understood, due to limitations of space this article will disregard the view. *IM* is to be distinguished from a more popular version of the simple view, according to which personal identity relations are weakly reductive (*WR*) and in independence non-informative (*INI*):

WR-INI: X at t_1 is identical to Y at t_2 iff there is some fact F_1 about X at t_1, and some fact F_2 about Y at t_2, and F_1 and F_2 are irreducible to facts about the subjects' psychology or physiology, and X at t_1 is identical with Y at t_2 in virtue of the fact that the propositions stating F_1 and F_2 differ only insofar as that "X" and "t_1" occur in the former where "Y" and "t_2" occur in the latter.

WR-INI is weakly reductive in the sense that, while the identity relation in question can be reduced to a further domain, the further domain itself typically exhibits elements of non-reducibility and/or resistance to full physical explanation. In their most prominent variants, these elements are due to references to souls, Cartesian Egos or other spiritual or immaterial substances and/or properties. Initially the idea underlying this claim may appear prejudicial; ultimately it is based on a number of widespread but not universally accepted beliefs about the naturalness of the world and the nature, validity and theoretical implications of physicalism.

According to this general stance, either both psychological and physiological continuity relations are fully reducible to a domain in which physical explanations are couched, perhaps in terms of the basic elements of a final and unified theory of physics, or they belong themselves to such a domain.

WR-INI may entail *IM* but does not so necessarily: it is conceivable that personal identity relations consist in something which is itself neither identical with nor reducible to a spiritual substance nor identical with nor reducible to aggregates or parts of psychologies and physiologies. In fact, Descartes' own view that personal identity is determined by "vital union" relations between pure Egos and bodies, with the persistence of the Ego being regarded as sufficient for the persistence of the person but the person not being wholly identifiable with the Ego, could be a weakly reductive view of persons. It is merely weakly reductive, however, because the identity of the phenomenon that specifies the necessary and sufficient conditions for personal identity does not itself follow from anything other than itself. While a weakly reductive criterion of personal identity relations is explicable in terms of the identities of phenomena other than persons, the identities of these phenomena themselves are not explicable in other terms: their identity may be, as we would suppose "soul identity" to be, "strict and philosophical", and not merely "loose and popular" (Butler 1736).

Nowadays, the Simple View is disparaged as a theory only maintained by thinkers whose religious or spiritual commitments outweigh the reasons that speak against their views on personal identity. This is due to the fact that it is assumed that a theory of personal identity cannot be weakly reductive without involving appeal to discredited spiritual substances or committing itself either to the acknowledgment of yet unrecognized physical entities or to an Identity Mysticism on the level of persons. As a consequence, many philosophers think that the problems that infiltrate dualism and Cartesian theories of the soul, such as the alleged impossibilities to circumscribe the ontological status of souls and to explain how a soul can interact with a body, render the Simple View equally problematic. Although the options mentioned are exceedingly difficult to defend, why should they have to be regarded as the only options available to the Simple Theorist? Arguably, many respectable philosophical ideologies, such as conceptualism or Neo-Kantianism, may issue in theories of personal identity along Simple lines without appeal to Cartesian Egos. (Note, however, that these ideologies, with regards to the problem of the persistence of people, may also be, and in fact have been, construed

along physiological or psychological lines). This suggests that we do not only need a better understanding, and above all more promising articulations, of the Simple View, but also a new taxonomy of theories of personal identity: the traditional division of theories into Simple, Psychological and Physical, even if maintained here by the author of this entry, may not be the best way of viewing the matter.

b. Reductionism (1): General Features

Modern day personal identity theory takes place mainly within reductionist assumptions, concentrating on the relative merits of different criteria of identity and related methodological questions. Reductionist theories of personal identity share the contention that…

Reduction: Facts about personal identity stand in an adequate reduction-relation to sets of sub-personal facts $SF_1 \ldots SF_n$ about psychological and/or physiological continuities in such a way as to issue in biconditionals of the form "X at t_1 is identical to Y at t_2 iff X at t_1 and Y at t_2 stand in a continuity-relation fully describable by SF_x."

Thus, any given set of sub-personal facts will impose demands, in forms of necessary and sufficient conditions, upon the kinds of adventures a subject can survive in persisting from t_1 to t_2. The sets of necessary and sufficient conditions determined by these sets of sub-personal facts constitute the various criteria of personal identity. It must be noted that the biconditionals in question need not to be understood in such a way as that circularity is an objection to them: provided that concepts other than "person" feature in the *analysans*, these biconditionals, by exhibiting connections with collateral and independently intelligible concepts, may be genuinely elucidatory even if the concept to be analyzed features on both sides of the equation (cf. McDowell 1997; Wittgenstein 1922, 3.263).

Only when the concepts "person" and "personal identity" become the target of what may be referred to as an *authentic* reduction circularities become vicious. The need for the distinction between authentic and inauthentic reductions arises due to an equivocation that ought not to confuse the present discussion: reductionisms in personal identity theory often take forms, if regarded for example as sets of supervenience claims, that are deemed, in other areas of analytic philosophy, as distinctively non-reductionist. Let us speak of authentic reductions if the ontological status of members of the reduced category is, in a way to be made precise, diminished in favor of the allegedly "more fundamental" existence-status of members of the

reducing category. The question of whether an authentic reductionism about persons must claim that it is not only able to give a criterion of personal identity without presupposing personal identity but also that facts about persons are describable without using the concept "person" is a matter of current controversy (cf. Behrendt 2003; Cassam 1989; 1992; Johnston 1997; McDowell 1997; Parfit 1984; 1999; forthcoming; cf. also 2.d.).

In a search for the necessary and sufficient conditions for the sustenance of personal identity relations between subjects, which type of continuity-relations could *SF* describe? There are two main contenders, physiological continuity-relations and psychological continuity-relations, which will be discussed in turn.

c. Reductionism (2): Psychological Approaches

Psychological Criteria of personal identity hold that psychological continuity relations, that is, overlapping chains of direct psychological connections, as those causal and cognitive connections between beliefs, desires, intentions, experiential memories, character traits and so forth, constitute personal identity (cf. Locke 1689, II.xxvii.9-29; Parfit 1971a; 1984; Perry 1972; Shoemaker 1970; Shoemaker & Swinburne 1984).

Two apparently physiological theories of personal identity are at bottom psychological, namely (i) the *Brain Criterion*, which holds that the spatiotemporal continuity of a single functioning brain constitutes personal identity; and (ii) the *Physical Criterion*, which holds that, necessarily, the spatiotemporal continuity of that which sustains the continuous psychological life of a human being over time, which is, contingently, a sufficient part of the brain that must remain in order to be the brain of a living person, constitutes personal identity (cf. Nagel 1971). These approaches are at bottom psychological because they single out, as the constituting factors of personal identity, the psychological continuity of the subject. Consider a test case. Imagine there to be a tribe of beings who are in all respects like human beings, except for the fact that their brains and livers have swapped bodily functions: their brains regulate, synthesize, store, secrete, transform, and break down many different substances in the body, while their livers are responsible for their cognitive capacities, basic integrated postural and locomotor movement sequences, perception, instincts, emotions, thinking, and other integrative activities. Imagine the brain criterion to be true for human beings. Would we have sufficient reason to believe the brain criterion to be true for members of the tribe in question as well, if we were aware of all facts about their physiologies? No, *precisely because* the brain criterion is true for human beings, a liver criterion

would have to be true for members of this tribe. There is nothing special about the 1.3 kilograms of grey mass that we carry around in our skulls, except for the fact that this mass is the seat of our cognitive capacities.

We can further distinguish between three versions of the psychological criterion: the *Narrow* version demands psychological continuity to be caused "normally," the *Wide* version permits any *reliable* cause, and the *Widest* version allows *any* cause to be sufficient to secure psychological continuity (cf. Parfit 1984). The Narrow version, we may note, is logically equivalent to the Physical Criterion.

One might think that brain criterion and physical criterion, to varying degrees, combine the best of both worlds: both acknowledge the vital function psychological continuity plays in our identity judgments while at the same time admitting of the importance of physiological instantiation. In fact, however, the opposite is the case: the appeal to physiology introduces an unacceptable element of contingency into the answers to the persistence question envisaged by defenders of these criteria. A criterion of personal identity tells us what our persistence *necessarily* consists in, which means that it must be able to deliver a verdict in possible scenarios that is consistent with its verdicts in ordinary cases. One scenario that has been widely debated is the following:

Teletransportation

At t_1, X enters a teletransporter, which, before destroying X, creates an exact blueprint of X's physical and psychological states. The information is sent to a replicator device on Mars, which at t_2 creates a qualitatively identical duplicate, Y (cf. Parfit 1984). *Our alleged intuition*: since Y at t_2 shares with X at t_1 all memories, character traits, and other psychological characteristics, X and Y are identical. *Alleged conclusion*: should teletransportation be reliable, all proposed criteria but the Wide and Widest versions of the Psychological Criterion are false.

Should teletransportation be unreliable, all criteria of personal identity but the Widest version of the Psychological Criterion are false. Consequently, should appeal to such scenarios as *Teletransportation* be acceptable and should the intuition above be widely shared, the brain criterion and physical criterion are false.

d. Quasi-Psychology

Many people regard the idea that our persistence is intrinsically related to our psychology as obvious. The problem of cashing out this conviction in theoretical terms, however, is notoriously

difficult. Psychological continuity relations are to be understood in terms of overlapping chains of direct psychological connections, that is, those causal and cognitive connections between beliefs, desires, intentions, experiential memories, character traits and so forth. This statement avoids two obvious problems.

First, some attempts to cash out personal identity relations in psychological terms appeal exclusively to *direct* psychological connections. These accounts face the problem that identity is a transitive relation (see 1.a.) while many psychological connections are not. Take memory as an example: suppose that Paul broke the neighbor's window as a kid, an incident he remembers vividly when he starts working as a primary school teacher in his late 20s. As an old man, Paul remembers his early years as a teacher, but has forgotten ever having broken the neighbor's window. Assume, for *reductio*, that personal identity consists in direct memory connections. In that case the kid is identical with the primary school teacher and the primary school teacher is identical with the old man; the old man, however, is not identical with the kid. Since this conclusion violates the transitivity of identity (which states that if an X is identical with a Y, and the Y is identical with a Z, then the X must be identical with the Z), personal identity relations cannot consist in direct memory connections. Appeal to overlapping layers or chains of psychological connections avoids the problem by permitting indirect relations: according to this view, the old man is identical with the kid precisely because they are related to each other by those causal and cognitive relations that connect kid and teacher and teacher and old man.

Second, memory alone is not necessary for personal identity, as lack of memory through periods of sleep or coma do not obliterate one's survival of these states. Appeal to causal and cognitive connections which relate not only memory but other psychological aspects is sufficient to eradicate the problem. Let us say that we are dealing with psychological *connectedness* if the relations in question are direct causal or cognitive relations, and that we are dealing with psychological *continuity* if overlapping layers of psychological connections are appealed to (cf. Parfit 1984).

One of the main problems a psychological approach faces is overcoming an alleged circularity associated with explicating personal identity relations in terms of psychological notions. Consider memory as an example. It seems that if John remembers having repaired the bike, then it is necessarily the case that John repaired the bike: saying that a person remembers having

carried out an action which the person did not in fact carry out may be regarded as a misapplication of the verb "to remember." To be sure, one can remember *that* an action was carried out by somebody else; it seems to be a matter of necessity, however, that one can only have first-person memories of experiences one had or actions one carried out. Consequently, the objection goes, if memory and other psychological predicates are not impartial with regards to identity judgments, a theory that involves these predicates and that at the same time proposes to explicate such identity judgments is straightforwardly circular: it plainly assumes what it intends to prove.

To make things clearer, consider the case of *Teletransportation* above: if at t_2 Y on Mars remembers having had at t_1 X's experience on earth that the coffee is too hot, then, necessarily, X at t_1 is identical with Y at t_2. The dialectic of such thought experiments, however, requires that a description of the scenario is possible that does not presuppose the identity of the participants in question. We would wish to say that since X and Y share all psychological features, it is reasonable or intuitive to judge that X and Y are identical, and precisely *not* that since we describe the case as one in which there is a continuity between X's and Y's psychologies, X and Y are necessarily identical. If some psychological predicates presuppose personal identity in this way, an account of personal identity which constitutively appeals to such predicates is viciously circular.

In response, defenders of the psychological approach have created psychological concepts that share with our ordinary psychological predicates all features except presumptions of personal identity: for example, the concept of "quasi-memory" is exactly like ordinary memory apart from the fact that "memory" is judgmental with regards to personal identity whereas "quasi-memory" is not (cf. Shoemaker 1970). While many commentators regard the appeal to quasi-memory, and ultimately "quasi-psychology," as sufficient to solve the circularity problem, some commentators think that personal concepts infiltrate extensionally articulated psychological concept-systems so deeply that any reductionist programme in personal identity is doomed from the start (cf. Evans 1982; McDowell 1997).

e. Reductionism (3): Physiological Approaches

Opponents of the psychological criterion typically favour a physiological approach. There are at least two of them: (i) the *Bodily Criterion* holds that the spatiotemporal continuity of a

functioning human body constitutes personal identity (cf. Williams 1956-7; 1970; Thompson 1997); and (ii) the *Somatic Criterion* holds that the spatiotemporal continuity of the metabolic and other life-sustaining organs of a functioning human animal constitutes personal identity (cf. Mackie 1999; Olson 1997a; 1997b; Snowdon 1991; 1995; 1996). It is not obvious that there is a straightforward relation between them, for everything depends on how the notions of "functioning human body" and "life-sustaining organs" are understood. If these notions are understood similarly, the views are (close to) equivalent; the other extreme, even if unlikely to be held, is that the notions are understood differently, to the effect that they are incompatible (if, for example, a functioning human body and its life-sustaining organs could come apart). Physiological approaches have consequences many of us feel uncomfortable with. Consider the following thought experiment:

Body Swap

X's brain is transplanted into Y's body. X's body and Y's brain are destroyed, the resulting person is Z. *Our alleged intuition*: since Z shares with X all memories, character traits, and other psychological characteristics, X is identical with Z. *Alleged conclusion*: the Bodily and the Somatic Criteria are false (cf. Locke 1689, II.xxvii.15; Shoemaker 1963).

Defenders of bodily criterion and somatic criterion typically bite the bullet and argue that it is not the case that X and Y have swapped bodies, but that Y falsely believes to be X, and therefore that Z is identical with Y.

Since the psychological and physiological approaches are mutually exclusive and, we may suppose in the current context, as candidates for an adequate theory of personal identity jointly exhaustive, any objection against the psychological approach is equally an argument for the physiological approach. The initial implausibility of the physiological approach is due to thought experiments that traditionally permeate the personal identity debate and often favour psychological considerations. Defenders of the somatic approach, most notably Olson and Snowdon, have tried to shift the focus to real-life cases in which descriptions along physiological lines look much more promising. Consider:

Human Vegetable

X has at t_1 a motor bicycle accident. The being Y that is transported to the hospital is at t_2 in a persistent vegetative state. *Our alleged intuition*: X at t_1 is identical with Y at t_2. *Alleged conclusion*: all views which postulate psychological continuity as a necessary condition are false.

Fetus

Since a fetus does not possess the cognitive capacities necessary to satisfy the demands of the Psychological Criterion, if the latter is true, no person can be identical with a past fetus. *Our alleged intuition*: Each of us is identical with a past fetus. *Alleged conclusion*: all views which postulate psychological continuity as a necessary condition are false.

A third problem for the psychological approach is that it implies, supposedly, that we are not human animals (Ayers 1990; Snowdon 1990; Olson 1997a; 2002a). The argument is simple:

Premise 1: Psychological continuity is neither necessary nor sufficient for the persistence of a human animal.

Premise 2: The psychological approach claims that psychological continuity is necessary and/or sufficient for our persistence.

A: for *reductio:*The psychological approach is true.
B: from 2, A: Psychological continuity is necessary and/or sufficient for our persistence.

Premise 3: Psychological continuity cannot at the same time be (i) necessary and/or sufficient for a thing's persistence and (ii) neither necessary nor sufficient for the same thing's persistence.

C: from 1, B, 3: None of us is identical with a human animal.

Premise 2 is implied by the psychological approach. The thought experiments that support premise 1 have already been given: since the human animal each of us is has been a fetus and could end up as a human vegetable, the thought experiments *Fetus* and *Human Vegetable* above demonstrate that psychological continuity is not necessary for human animal identity. A variant of *Body Swap* shows that psychological continuity is not sufficient for human animal identity. Suppose X's brain to be transplanted into Y's skull and X's body and Y's brain are destroyed.

Suppose further that the resulting being Z is psychologically continuous with X. In this case, it does not seem to be the case that the surgeons transplant the human animal X from one head to another. Rather, it seems, the human animal Y receives a new organ, namely a brain. Consequently, psychological continuity is not sufficient for human animal identity and premise 1 holds. Premise 3 seems to be obvious, because its being false would entail that one and the same being can outlive itself, which is absurd. The defender of the physiological approach now argues that

Premise 4: We are human animals.

C: from B, 4: The psychological approach is false.

Premise 5: Physiological and psychological answers to the persistence question are mutually exclusive and jointly exhaustive.

Conclusion: The physiological approach is true.

It may be argued that premise 4 is not a matter of metaphysics but of biological classification. The underlying problem, however, is that it seems undeniable that there is a human animal located where each of us is. If this human animal has persistence conditions different from those that determine *our* persistence, then there must be two things wherever each of us is located. This conclusion raises important questions and problems a psychological approach must address.

3. The Paradox of Personal Identity

One of the most influential thought experiments in recent personal identity theory is the case of fission.

a. Fission

Fission

X's brain is removed from X's body and X's body is destroyed. X's brain's corpus callosum, the bundle of fibres responsible for retaining the capacity of information-transfer between the two brain hemispheres, is severed, leaving two (potentially) equipollent brain hemispheres. The single lower brain is divided and each hemisphere is transplanted into one of two qualitatively identical bodies of the fission outcomes Y_1 and Y_2. *Our alleged intuition*: since both Y_1 and

Y_2 share with X all psychological characteristics, both are candidates for being identical with X: either, in the absence of the other, would have been identical with X. *Alleged conclusion*: either, on pain of violating the transitivity of identity, the Psychological Criterion is false or the question of whether two person-stages X at t_1 and Y_1 at t_2 are temporal parts of the same person depends on facts concerning not only X and Y_1 but also, in this case, Y_2. In the latter case, a "closest continuer" clause and/or a "no-branching" proviso must complement a psychological continuity analysis (For a development of this case, see Nozick 1981; Parfit 1984; and Wiggins 1967).

Fission scenarios emphasise the difficulty of deciding whether a thought experiment is acceptable or not. They assume the possibility of commissurotomy or brain bisection, that is, the perforation of the corpus callosum, and hemispherectomy, that is, the surgical removal of the cerebral cortex of one brain hemisphere. Commissurotomy was used in epilepsy treatment in the 50's (cf. Nagel 1971) and hemispherectomies too have been performed in the past. However, fission cases additionally assume the possibility, in some sense or other, of dividing the subcortical regions, and in particular the single lower brain. This is not physically possible without damaging the upper brain functions (cf. Parfit 1984). Many commentators regard fission to be an acceptable challenge to theories of personal identity. Wilkes disagrees: she thinks that our ignorance about what actually happens in these cases jeopardises the theoretical relevance of fission scenarios (cf. 1988). The question of whether or not physically impossible but logically possible scenarios are acceptable remains to be answered.

Should fission be an acceptable scenario, it presents problems for the the psychological approach in particular. The fission outcomes Y_1 and Y_2 are both psychologically continuous with X. According to the psychological approach, therefore, they are both identical with X. By congruence, however, they are not identical with each other: Y_1 and Y_2 share many properties, but even at the very time the fission operation is completed differ with regards to others, such as spatio-temporal location. Consequently, fission cases seem to show that the psychological approach entails that a thing could be identical with two non-identical things, which of course violates the transitivity of identity. Some commentators have attempted to save the psychological approach by appeal to the so-called "multiple occupancy view," that is, the claim that, despite appearances, X was two people, namely Y_1 and Y_2, all along (cf. Lewis 1976; Noonan 1989;

Perry 1972). Combined with a four-dimensionalist or temporal part ontology, this view is not as absurd as it initially seems, but it is certainly controversial.

Others have acknowledged, as a consequence of fission scenarios, that psychological continuity is not sufficient for personal identity. These commentators typically complement their psychological theory with a non-branching proviso and/or a closest continuer clause. The former states that even though X would survive as Y_1 or Y_2 if the other did not exist, given that the other does exist, X ceases to exist. This proviso avoids the problem of violating the transitivity of identity. It is hard to believe, however, because it entails that I can kill you without you ever noticing: if I knock you unconscious, transplant one of your brain hemispheres into a different body, and drop you off at home before you wake up, then, if the transplant is successful and the psychological approach with non-branching proviso is true, you are dead. We could avoid this problem by adding a closest-continuer or best candidate clause, stating roughly that the best candidate for survival in a fission scenario, that is, the fission outcome which bears the most or the most important resemblances to the original person X, is identical with X. One of the problems with this suggestion is that it assumes that personal identity is an extrinsic relation. It thereby violates another important principle, namely the so-called "only X and Y rule," which states, roughly, that if two person-stages at different times are stages of one and the same person, that will be true only in virtue of the intrinsic relation between these two stages (cf. Noonan 1989; Wiggins 2001). While this principle is not necessarily sacrosanct, it is desirable to avoid violating it.

b. The Paradox

The upshot of the preceding discussion is that we find ourselves in a perplexing situation. Let the underlying assumption be that there is a criterion of personal identity. The starting point of the debate has been that

Premise 1: A criterion of personal identity captures all those aspects of our existence that are necessary and sufficient for our persistence.

Premise 2: Our persistence is determinate.

A: from 1, 2: A criterion of personal identity determines for every possible past event e_0 and future event e_2, within the boundaries of an adequate delineation of the modality in question,

whether a person X at t_1 is identical with the being that has participated in e_0 and the being that will participated in e_2.

Premise 3: Personal identity relations are *factual*: criteria of personal identity are determined neither by conventions, norms, or other social or personal preferences, however basic, nor by analytic matters about the meaning of concepts. Their truth is, literally, a matter of life and death.

B: from A, 3: There is a factual relation *R* between a person X at t_1 and a being Y at t_0/t_2 which, for every possible scenario, determines whether X at t_1 is identical with Y at t_0/t_2.

Now, if we agree with the tentative conclusion that there is, at present, no satisfactory simple view of personal identity, then we assent to the claims that

Premise 4: *IM* and *WR–INI* are, with respect to a specification of the necessary and sufficient conditions for personal identity, inadequate.

Premise 5: The distinction between *IM* and *WR–INI* on the one hand and the reductionist views sketched in I.A.4 on the other is exclusive.

C: from 4, 5: The only feasible candidates for *R* are relations of physiological and/or psychological continuity.

Since B demands that *R* holds for every possible scenario, within the limits of an adequate delineation of the modality in question, a criterion of personal identity must deliver compatible judgments on the thought experiments sketched above. However, since these thought experiments deliver conflicting intuitions about which criterion is true, it cannot be the case that more than one such criterion is true. From this it follows that

Premise 6: Physiological and psychological criteria of personal identity are incompatible, that is, *R* cannot be a conjunction of physiological and psychological relations *as well as* issuing in determinate and compatible solutions to each thought experiment.

Now, if we are also prepared to accept the

Big Assumption: A criterion of identity must accept all alleged conclusions of the thought experiments sketched in I.A.5

then we must conclude that

D: from B, 6A: Neither physiological nor psychological continuity is both necessary and sufficient for personal identity.

The problem with D is that, in conjunction with premises 2, 4, and 5, it reduces the underlying assumption that there can be an informative criterion of personal identity *ad absurdum*. This argument may be referred to as the *Paradox of Personal Identity*.

One should refrain from drawing precipitate conclusions from its defining characteristic as a paradox, that is, the fact that denying any of its premises leads to a conclusion that either violates our intuitions or, in the case of 4, 5, and C, commits one to a philosophically disreputable stance. Rather, the Paradox should be regarded as the starting point of any discussion of personal identity, in the sense that taking a stand on its individual premises bestows the various criteria of personal identity with their distinctive features. However, *given* that the paradox obliges us, in one way or other, to revise our pre-philosophical beliefs, a theory of personal identity should aim at meeting what will be referred to as the *Adequacy Constraint AC* on theories of personal identity, which demands that

AC: We ought to sanction a substantial revision of our pre-philosophical views of our metaphysical nature only on the conditions that (i) we receive an explanation of the unreliability of our intuiting faculties that in this domain outweighs our grounds for, and in other domains is compatible with, believing in their reliability; (ii) we receive an approximate demarcation of the extents to which we have to abandon our pre-philosophical beliefs and to which we can and we cannot have knowledge about ourselves.

How is the Paradox to be resolved? A, B, C, and D are deductions, and premise 1 is plausible on independent grounds. If identity is determinate, then premise 1 is true. Consequently, those arguments that deny the possibility of vague objects and indeterminate identity, in addition to our intuition that our own identity must be determinate, work in favor of 1. Note that, should personal identity be indeterminate, we might still be able to give a criterion of personal identity, even though such a criterion would then fall short of giving full necessary and sufficient conditions, since in some imaginary case it does not apply.

The denial of premise 3 seems to entail that we have, in a deep sense, an influence on whether we survive a given adventure, namely by possessing a particular normative, experiential, or

attitudinal background. This contention may contradict our intuitions more than any thought experiment could. Since we assumed premises 4 and 5, only premises 2 and 6 and the *Big Assumption* remain. Could one deny premise 6? Given that the determinacy and factuality premises are accepted, It is hard to believe that we could: if a hybrid view were determinately true, a human being could die twice, once when her psychological and once when her physiological capacities cease to function. As a result, most commentators accept 6 but choose to accept a particular criterion in the vicinity of either side of the psychology-physiology divide. This implies that the *Big Assumption* must either not entail D or be rejected, which can be argued, always assuming that *AC* is being met, in three ways:

(a) One could define "adequacy of modality" in such a way as to exclude precisely those thought experiments which are problematic for a given criterion. There are two problems with this proposal: first, it is difficult to see how such a definition of adequacy of modality could *not* be *ad hoc*. And secondly, the suggestion is insufficient, for some thought experiments circumscribing *physically possible* scenarios, such as *Human Vegetable*, trigger incompatible intuitions as well. While some commentators think that Y is identical with X despite X's loss of cognitive capacities, others regard Y as a living grave stone, nurtured merely for sentimental reasons, in commemoration of the deceased X.

(b) One could deny premise 2 instead, arguing that if personal identity is indeterminate, then our preferred criterion of personal identity does not have to deliver verdicts in all thought-experimental scenarios. This move has the further benefit that we do not have to quarrel with the alleged conclusion of another thought experiment, the combined spectrum:

Combined Spectrum

A spectrum of possible cases is imagined: at the near end, the normal case, X at t_1 is fully psychologically and physiologically continuous with Y at t_2, while at the far end X at t_1 is neither psychologically nor physiologically continuous with Y at t_2. In the intermediate cases, X at t_1 is approximately halfway psychologically and physiologically continuous with Y at t2. *Our alleged intuition*: towards the near end of the spectrum X at t_1 is identical with Y at t_2 and towards the far end of the spectrum X at t_1 is not identical with Y at t_2. There could not even in principle be evidence for the existence of a sharp borderline between the cases in which X at t_1 *is* and the

cases in which X at t_1 *is not* identical with Y at t_2. Hence, it is implausible to believe that such a borderline exists. *Alleged conclusion*: personal identity is indeterminate.

Epistemicists like Timothy Williamson (cf. 1994) deny that we should render it implausible that there is such a sharp borderline merely because we are necessarily ignorant of its existence. Vagueness, according to epistemicism, consists precisely in our necessary ignorance of such sharp boundaries. The other problem is that even if personal identity is indeterminate, the claim cannot *by itself* establish one criterion over others: in order to do so, it would have to exclude those thought experiments that challenge opposing criteria while leaving untouched those that supposedly establish the preferred criterion. It is doubtful, however, that the indeterminacy of personal identity can be exploited selectively, for physiological and psychological continuity relations are equally indeterminate in a particular range of cases (cf. Parfit 1984). Furthermore, in those cases in which they are not, for example *Body Swap, Human Vegetable,* and *Fetus,* appeal to indeterminacy does little to remove the contradictory intuitions that these cases trigger. Consequently, unless one holds that personal identity is categorically indeterminate whenever the physiological and psychological features of a human being come apart, appeal to indeterminacy cannot establish the rejection of the *Big Assumption* in such a way as to avoid the Paradox's conclusion.

(c) The most common strategy is to bite the bullet and some or other allegedly absurd conclusion of the thought experiments. The defender of the Psychological Criterion must hold that we are not identical with a past fetus or infant, and that we will not have survived if fallen into a persistent vegetative state. Defenders of a Physiological Criterion, on the other hand, must commit to the consequence that if X's head is grafted onto Y's body, then the resulting person is Y and not X, even though this person shares all psychological features with X before the operation.

The problem with this strategy is that, if accepted, we seem to be unable to decide on a criterion of personal identity on the basis of intuitions at all, on pain of unjustifiably favoring one's own over other people's intuitions. On the assumption that we are unable to hierarchically structure these conflicting intuitions, we have a classical stand-off: there are two sides to the coin of personal identity and appeal to intuition plainly underdetermines preferring one side over the other. The problem is that human beings are organic material objects, the persistence of which is

determined by these objects' following a continuous trajectory between space-time points. The further question of whether or not human beings are *essentially* organic material objects depends on the question of whether psychological properties render human beings to be *sufficiently dissimilar* from such objects so as to "deserve" their own identity criterion. The fear underlying the Paradox of Personal Identity, then, is that there may be no metaphysical fact to the matter as to whether the antecedently specifiable differences between human beings and other organic or inorganic material objects *count as sufficient* in order for us to have persistence conditions different from these objects. It does not seem as if any possible thought experiment, irrespectively of how unequivocal our intuitions about it, could redeem this fear. Personal identity theorists, therefore, ought to offer a more comprehensive account of the ontological status of persons and their relation to the constituents that make them up.

4. Parfit and the Unimportance of Personal Identity

Derek Parfit proposes a theory of the ontological status of persons, which promises to answer the problem of fission and the paradox of personal identity. While this article cannot do justice to the complexities of Parfit's theory, which has been the focal point of debate since 1970, it is worth mentioning its main features.

Although Parfit affirms the existence of persons, their special ontological status as non-separately-existing substances can be expressed by the claim that persons do not have to be listed separately on an inventory of what exists. In particular, persons themselves are distinct from their bodies and psychologies, but the existence of a person consists in nothing over and above the existence of a brain and body and the occurrence of an interrelated series of mental and physical events. These are the foundational claims of Parfit's constitutive reductionism. Consider an analogy: Cellini's *Venus* is made of bronze. Although the lump of bronze and the statue itself surely exist, these objects have different persistence conditions: if melted down, *Venus* ceases to exist while the lump of bronze does not. Therefore, they are not identical; rather, so the suggestion, the lump of bronze constitutes the statue. The same is true of persons, who are constituted by, but not identical with, a physiology, a psychology, and the occurrence of an interrelated series of causal and cognitive relations.

Now, how does this relate to the fission case? We must first note that Parfit believes (i) that our persistence consists in physical and/or psychological continuity; (ii) that personal identity is

indeterminate in some cases, that is, that sometimes there is no right-or-wrong answer to the question of whether somebody has ceased to exist in the course of a certain adventure (see 3.b.); (iii) that what prudentially matters in survival is psychological continuity; (iv) that personal identity relations must respect the remaining formal properties of identity. This means that in the fission case Y_1 and Y_2 cannot be identical with X because the transitivity of identity is violated: therefore, X dies in the fission case. It further means, however, that X has two Parfitian survivors, Y_1 and Y_2, which is, according to Parfit, as good (or even better) than being identical with Y_1 and/or Y_2. This is the upshot of Parfit's claim that what prudentially matters is psychological continuity: for all we should care, from a purely rational point of view, it is good enough for us to be psychologically continuous with one or more future persons and consequently it would be irrational for us to prefer our own continued existence to death by fission. Generally, according to Parfit, psychological continuity with any reliable cause matters in survival, and since personal identity does not consist merely in psychological continuity with any reliable cause, personal identity is not what matters in survival.