8

REASONS AND CONSCIOUS PERSONS
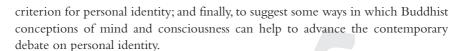
Christian Coseru

Introduction

What justifies holding the person that we are today morally responsible for some-thing we did a year ago? And why are we justified in showing prudential concern for the future welfare of the person we will be a year from now? Whatever our answer to these questions, it seems that we cannot systematically pursue them without in one way or another referring to persons and their identity over time. But while there is widespread agreement that considerations about personal identity must be front and center in any such inquiries, such agreement falls short when it comes to specifying the criteria for personal identity, that is, what this identity necessarily involves or consists in. Part of the difficulty is that an investigation into the nature of personal identity brings us to metaphysical questions about persons, their onto-logical status, identity conditions, and persistence over time. The challenge, then, is to pursue these additional questions without losing sight of the practical concerns that prompted them in the first place.

Few contemporary philosophers have confronted this challenge with more ana-lytic skill, depth, and ingenuity than Derek Parfit. In engaging with Parfit's work on personal identity, primarily his *Reason and Persons*, my aim is to reassess his Reductionist View of personal identity in light of Buddhist Reductionism, a philo-sophical project grounded on the idea that persons reduce to a set of bodily, sen-sory, perceptual, dispositional, and conscious elements, which alone are real. Parfit is not only familiar with this Buddhist conception of personal identity, but thinks that the reductionist, no ownership position he defends, which takes persons both to exist and to reduce to their components, is true, and that, as he famously puts it, "Buddha would have agreed" (1984: 273). My goal here is threefold: first, to review Parfit's Reductionism position and evaluate its main arguments; second, to assess the extent to which Buddhist Reductionism supports Parfit's psychological

criterion for personal identity; and finally, to suggest some ways in which Buddhist conceptions of mind and consciousness can help to advance the contemporary debate on personal identity.

Reductionism and personal identity

What is reductionism and how does it relate to the question of personal identity? The literal sense of the term, derived from the Latin '*reducere*' (lit. 'bring back', 'lead back'), captures rather well the philosophical notion of reduction: to reduce is 'to lead back' to something more fundamental than the thing in question. To say that the mental reduces to the physical, that motion reduces to kinetic energy, and that psychology reduces to biology, is to capture what it means for one thing (mind, motion, or psychology) to be brought back to the other (physical, kinetic energy, or biology). Reduction, then, stands for the view that, if entity x reduces to entity y, then y is prior to, more basic than, or constitutive of x.[1] As a philosophical term of art, 'reduction' is central both to metaphysical questions of personal identity and to questions in philosophy of mind about consciousness, agency, and the mind-body problem. One is a reductionist if one takes a particular theory or phenomenon to be conceivable in terms of, or reducible to, another theory or phenomenon.[2]

Parfit is one such reductionist, and *Reasons and Persons* contains his vastly influential theory of personal identity, which argues against the commonsensical, non-reductionist view of persons. According to the non-reductionist view, persons are distinct and discrete entities that exist over and above their bodies and psychological states. Their identity, then, is an irreducible, brute fact of existence, and cannot be explained or described in more basic terms. Whatever persons are, an account of their identity would have to employ person-level descriptive categories of embodied experience.[3] One paradigmatic example for person in this non-reductive sense is the Cartesian Ego. For Parfit the view that there are such entities as Cartesian Egos or souls is representative of a particular intuition about personal identity, according to which we assume that questions of the sort 'Will I survive the death of my body?' or 'Will I be the same person if I were to be teleported elsewhere?' must have definitive answers. Regardless of whether or not we have answers to these questions at present, given their implications for personal identity, answers must in principle be available. There must be a way to settle these questions one way or another, perhaps on the basis of our very conception of what personal identity entails. What drives this intuition, argues Parfit, is the assumption that our identity must in some sense be *determinate* (1984: 214).

If we reject this intuition and allow for the possibility that Cartesian Egos do not exist, then, argues Parfit, we are in a sense compelled to accept the view of Reductionism. One of the advantages of Reductionism is that it offers new possibilities for reconceiving the problem of personal identity on both metaphysical and empirical grounds. Since the body is the seat of our physical, affective, and mental lives, we may conceive of persons as (i) bodies or as (ii) entities that have bodies, thoughts, and emotions. The first conception can also be understood as an
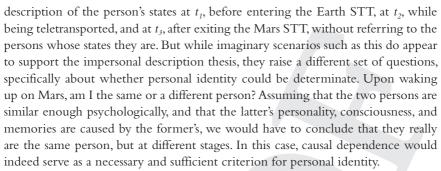
endorsement of one version of the *identity* view (persons just *are* bodies), while the second makes the case for the *ownership* view (persons are the sort of entities that *have* bodies, thoughts, and other kinds of experiences). As Parfit makes clear, whereas the ownership or constitutive view of personal identity can be easily entertained, and may even fit classical conceptions of persons as property-possessors, the identity view of reductionism opens the door for something more radical: Eliminative Reductionism, the best example of which is the mind/brain identity theory. For the eliminative reductionist, persons, much like nations, are ultimately reducible to their constitutive parts. Just as the concept of a nation cannot be the concept of an entity that is distinct from its people and its territory, so also the concept of a person cannot be that of an entity that exists over and above its body, perceptions, thoughts, and feelings. We may thus conclude, argues Parfit, "that, in that case, there are really no such things as nations. There are only groups of people, living together in certain ways" (2002: 656).

What does it mean to extend this analogy and to claim that there are really no such things as persons? On Parfit's general characterization of reductionism, the existence of persons, then, "just involves the existence of a brain and body, the doing of certain deeds, the thinking of certain thoughts, the occurrence of certain experiences" (1984: 2011). The force of this reductionist move is captured by the 'just consists in' locution: all of the facts that pertain to a person's existence—their particular type of body, their psychological profile, the kinds of relations that obtain between their mental and brain states, all of these admit of an impersonal description that neither requires nor presupposes this particular person exist. As a *façon de parler*, however, we do say that, because we can ascribe thoughts to them, thinkers exist. But just because we can infer from the content of their experience that there are thinkers, it does not follow that their existence is somehow separate from their thoughts. Parfit endorses the impersonal description thesis (1984: 225) even as he acknowledges, with Peter Strawson and Bernard Williams, that any talk of experiences, and of the various relations that obtain between them, is impossible without reference to the persons whose experiences they are. Although he would subsequently disavow it (Parfit 1999), on the grounds that a thoroughly impersonal conceptual schema would be ineffective for the kind of creature we are, the impersonal description thesis plays an important explanatory function for the view that personal identity over time just consists in physical and/or psychological continuity.

To motivate both reductionism and the impersonal description thesis, Parfit provides the well-known example of the simple teletransportation (STT) machine. Suppose that in a future century when science is far advanced I am able to travel to a distant place, say Mars, simply by having a machine scan every cell of my body and broadcast the information in real time to another machine on Mars. There is one catch, however: my current body will be destroyed in the process of being scanned. On the up side, a qualitatively identical but numerically different body will be created on Mars. The initiation of the whole process is marked by a loss of consciousness, which is immediately gained once the complete body has been recreated on Mars. In principle, argues Parfit, we should be able to provide a

description of the person's states at t_1, before entering the Earth STT, at t_2, while being teletransported, and at t_3, after exiting the Mars STT, without referring to the persons whose states they are. But while imaginary scenarios such as this do appear to support the impersonal description thesis, they raise a different set of questions, specifically about whether personal identity could be determinate. Upon waking up on Mars, am I the same or a different person? Assuming that the two persons are similar enough psychologically, and that the latter's personality, consciousness, and memories are caused by the former's, we would have to conclude that they really are the same person, but at different stages. In this case, causal dependence would indeed serve as a necessary and sufficient criterion for personal identity.

But as Parfit concedes, such scenarios are not entirely unproblematic for the psychological theory of personal identity. Consider the second scenario, where an advanced teletransportation (ATT) machine has the same scanning capabilities as the STT, plus the added feature of leaving the scanned body intact. In this case, having undergone teletransportation, and on the assumption that the person I am on Earth is psychologically very similar to the one I am now on Mars, one would find oneself in two places at once. One advantage of this second scenario is that it provides a compelling reason for adopting some version of the 'body' theory of personal identity. But neither scenario addresses the many objections to the impersonal description thesis such as, for instance, that reductionism cannot explain in a non-question-begging way how exactly we get persons from impersonal psycho-physical elements.

The teleportation examples aim to drive home the point that personal identity is indeterminate, and that reductionists must contend with a situation in which questions about the mode of existence of persons may yield no definitive answers.[4] Notwithstanding these problems, Parfit thinks such examples are compelling enough to warrant that we take reductionism seriously, and presents us with two distinct conceptions of reductionism about persons:

(1) a person just is a particular brain and body and a series of interrelated physical and mental events; and
(2) a person is an entity that is *distinct* from a brain and body and such a series of events; (1984: 211)

Although he thinks that (1), which articulates the eliminativist version of reductionism, is justified in some instances (it is right to claim that "there were really no witches, only persecuted women" (2002: 657)), his conception of persons as rational and moral agents clearly compels him to accept (2). Furthermore, Parfit insists on the importance of thinking about persons not in terms of their mental states, which would entail that they must be the states of some entity, but in terms of the occurrence of certain experiences, deeds, and thoughts. Thus understood, all reductionists would accept the view that:

(3) a person's existence just consists in the existence of a brain and body, and the occurrence of a series of interrelated physical and mental events (1984: 211).

Now contrast this view with that of the non-reductionist:

(4) a person is a separate existing entity, distinct from their brain and body, and their experiences (1984: 210)

But Parfit clearly distinguishes between two non-reductionist views. The common one, already noted, is that of a person as a purely mental entity, a Cartesian Pure Ego, soul, or spiritual substance. On this view, "personal identity over time does not just consist in physical and/or psychological continuity" (1984: 210). Rather, the Cartesian Ego is an altogether different metaphysical entity. The second non-reductionist view denies that persons are separately existing entities, distinct from their brains, bodies, and experiences:

(5) though we are not separately existing entities, personal identity is a further fact, which does not just consist in physical and/or psychological continuity (1984: 210)

Parfit calls this the *Further Fact View*. The typical non-reductionist view articulates a familiar conception of personal identity, whose basic premise is that human beings (and perhaps other forms of life) must be conceived in terms of something essential, without which we could not ascribe to them the properties of persons. Unlike the typical non-reductionist view, the further fact view poses a conundrum: what we have here is both the notion that persons are not metaphysically distinct entities and the notion that their existence cannot be accounted for in purely impersonal terms. On the further fact view, then, the concept of person is simple and unanalyzable. Parfit's account of personal identity, which had initially drawn on the reductionist no-self view of Buddhist Abhidharma,[5] may be more profitably associated with one Buddhist school of thought, Personalism (Pudgalavāda), that did indeed defend a conception of person as irreducible (more about the arguments adduced in its support in the next section).

Before we turn to the Buddhist discussions of personal identity, a brief overview of Parfit's psychological criterion of personal identity is in order.[6] Parfit takes the 'criterion' for identity to stand for "what this identity necessarily involves, or consists in" (1984: 202). In the case of physical objects such as the pyramids or the Moon that simply refers to the concrete spatio-temporal continuity of the object in question. So far as our physical body is concerned, the criterion of personal identity would simply consist in the persistence of this body and brain over time. Parfit offers what he assumes is a better version of the '*physical criterion*': what is necessary for personal identity is the persistence of *enough* of the body and the brain over time, such that even if we were to lose some of the body and brain (through amputation or hemispherectomy) the same person would continue to exist. Nonetheless, on the physical criterion of personal identity, the case of STT would not count as persistence over time and neither would that of rebirth, which presuppose that something of oneself continues beyond the destruction of the body. This is largely

the reason why Parfit thinks the physical criterion does not suffice, and why instead he puts forward a psychological criterion of personal identity, which he understands as involving two relations:

> *Psychological connectedness* is the holding of particular direct psychological connections.
> *Psychological continuity* is the holding of overlapping chains of strong connectedness

<div align="right">

1984: 206

</div>

Although Parfit takes connectedness, which is a matter of degree, to be a more important relation, he does not think it serves as an adequate criterion for personal identity. Strong connectedness is a transitive relation, since the person I am today is strongly connected to the one I was yesterday, and to the one I was three days ago. But it cannot be true that the person I am today is strongly connected to the one I was twenty years ago, and so, despite its importance, the strong relation of psychological connectedness is too problematic for personal identity: "Because identity is a transitive relation, the criterion of identity must also be a transitive relation. Since strong connectedness is not transitive, it cannot be the criterion of identity" (1984: 206). What about psychological continuity? Parfit uses the continuity relation to introduce his '*psychological criterion*' of personal identity:

> There is *psychological continuity* if and only if there are overlapping chains of strong connectedness. X today is one and the same person as Y at some past time if and only if (2) X is psychologically continuous with Y, (3) this continuity has the right kind of cause, and (4) it has not taken a 'branching' form. (5) Personal identity over time just consists in the holding of facts like (2) to (4).

<div align="right">

1984: 206

</div>

The key elements here are (2) psychological continuity and (3) right causal determination, and Parfit uses right causal determination to further distinguish between *narrow*, *wide*, and *widest* versions of the psychological criterion depending on whether the cause is identifiably the correct one, any reliable cause, or any cause whatsoever. These additional criteria provide a basis for introducing the memory element, since being able to remember past experiences as one's own is essential to the psychological criterion of personal identity. Memory thus stands as an example of strong connectedness between, for instance, an experience once had and later remembered or an intention once formed and later acted upon. These sorts of direct causal connections that memory, intention, and action provide are typically of a short duration. I seldom act on intentions formed twenty years ago, and distant childhood memories have already been embedded in a complex self-narrative that had undergone revision and embellishment over time (even though Parfit takes it to be a logical truth that we can only remember our own experiences).[7] So, while

there may be few direct connections between, say, the 50-year-old adult and the 10-year-old child, they are still indirectly connected through multiple overlapping chains of direct connectedness. The overlapping chains model thus serves to illustrate Parfit's understanding of the criterion for psychological continuity, whose functionality is such that we can talk of the child I was once and the adult I am today in one breath.

While the narrow account of causal determination serves as a plausible way to understand persistence over time, it cannot accommodate the simple teletransportation case. For this reason, Parfit thinks we need the wide and even widest views of causality. The wide view allows any reliable causal chain to serve as a criterion for personal identity. The widest view is even more permissible, allowing for any causal chain to play this role (a requirement for the persistence of identity in the ATT case). Since Parfit claims that there is no good reason to prefer the narrow view over the others, we are confronted with a problematic notion of persistence: just what it means to say the person I am now on Mars is psychological continuous with the one I was on Earth an hour ago is not exactly clear. And it is even more puzzling to entertain what it would be like to talk of personal identity between the two versions of myself in the ATT case. What is clear is Parfit's motivation for articulating these two views, both of which allow for a conception of persistence over time that reduces the person to certain chains of causal connectedness between mental and physical states. The issue is not whether the same individual persists over time, including in the ATT case. Rather, the problem is how best to guarantee there is enough resemblance between the different person stages. Even so, in articulating his psychological criterion, Parfit does not aim to secure a strong foundation for personal identity; rather, he wants to demonstrate that the persistence of identity is not an important factor when it comes to matters of moral and practical concern.

Now that we have identified the two primary elements of Parfit's theory of personal identity, let us consider to what extent Buddhist Reductionism supports Parfit's psychological criterion for personal identity.

Buddhist reductionism, personalism, and the no-self view

As is well known, in an effort to demonstrate that his views are not the product of a particular culture and epoch, but rather apply to all people at all time, Parfit turns to, inter alia, Buddhism, where he finds an early reductionist stance not unlike his own. As he notes:

> I claim that, when we ask what persons are, and how they continue to exist, the fundamental question is a choice between two views. On one view, we are separately existing entities, distinct from our brain and bodies and our experiences…The other view is the Reductionist View. And I claim that, of these, the second view is true… *Buddha would have agreed*.

> *1984: 273*

Whether or not Buddha would have agreed depends in large measure on whether
Buddhist Reductionism articulates something close to a psychological criterion of
personal identity. I will argue that it does, although in a way that emphasizes the
structural dynamics of conscious experience over accounts of psychophysical causal
relations (of the sort that Parfit's reductionist view of persons endorses). One of
the clearest articulations of Buddhist Reductionism about persons is found in *The
Questions of Milinda* (*Milindapañha*), an eloquent para-canonical work in the style of
a Platonic dialogue detailing an exchange between Nāgasena (a Buddhist monk)
and Milinda (better known as the Bactrian King Menander I) aimed at articulating
a nominalist conception of personal identity. In a much-quoted passage, Nāgasena
explains that, although he responds to a certain appellation—his name—there is no
corresponding referent:

> Sir, I am known as 'Nāgasena'; my fellows in the religious life address me
> as 'Nāgasena'. Although my parents gave (me) the name 'Nāgasena'…it is
> just an appellation, a form of speech, a description, a conventional usage;
> 'Nāgasena' is only a name, for no person is found here.
>
> *Horner 1963: 34*

Parfit is familiar with *The Questions of Milinda*, and appeals to this and to several
other excerpts from Buddhist texts[8] to make the case that "the Reductionist View
is not merely part of one cultural tradition. It may be, as I have claimed, the true
view about all people at all times" (1984: 273). That Parfit's view and that of the
Buddhists bear a strong resemblance is both obvious and well documented.[9] What is
less obvious is the extent to which Buddhist metaphysics supports the Reductionist
View. To answer this question, I will turn to one of the most influential attempts to
show how Buddhist resources could help Parfit address some of the difficulties of his
reductionist account while working out the implications of a Parfitian-style concep-
tual framework for Buddhist philosophy in general, and for Buddhist Reductionism
in particular: Siderits' *Personal Identity and Buddhist Philosophy* (1997/2015).[10]

Following Parfit, Siderits too acknowledges that reductionism comes in a var-
iety of forms, and that when situated in the Buddhist context one must not over-
look the (historical) fact that not all Buddhists are reductionists. Nonetheless,
within the context of early Buddhist scholasticism (viz., Abhidharma), all Buddhist
philosophers who, in addition to being reductionists about composite entities such
as chariots, forests, and armies, also take a reductionist stance with regard to persons,
may be identified as Buddhist Reductionists. As Siderits explains, a distinctive fea-
ture of Buddhist Reductionism is the two truths framework, which takes sentences
to be either conventionally or ultimately true depending on whether they lead to
successful practice or to what is taken to be ultimately real (2015: 16). For instance,
Buddhists take partite entities such as chariots to be conceptually constructed and
thus not ultimately real. Thus chariots reduce to 'axle', 'felly', and 'linchpin', which
in turn reduce to their elemental parts, and so on. On this account, only impartite
unanalyzable entities are ultimately real. Siderits reworks the two truths framework

into an account of semantic insulation, which he uses to justify the Buddhist Reductionist view that partite entities cannot be ultimately real. As he himself admits, this view, which articulates the position of mereological nihilism, "will strike many as giving an implausibly austere ontology" (2015: 14). Siderits finds some of the arguments adduced in its support compelling, though, as I will argue, for reasons that are not entirely clear. What, then, motivates Buddhist Reductionists' mereological nihilism? The answer lies in the widely-shared but, according to the reductionist, obviously mistaken non-reductionist stance that most people adopt:

> When the Buddhist Reductionist claims that strictly speaking chariots do not exist, that 'chariot' is a mere conventional designator for a collection of part arranged chariot-wise, this is meant to serve as an example of widely held belief that turns out not to be strictly speaking true, but also of a belief whose acceptance is perfectly understandable given the demands of everyday life. The point is not to get us to stop believing there are chariots. The point is to help us see how we could all be mistaken about 'I'. For the mechanism that generates belief in a real 'I' is the same; hypostatization. What we have in both cases is a many masquerading as a one.
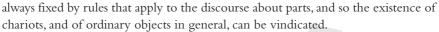>
> *Siderits 2015: 98*

Does use of the first-personal pronoun and handling of the middle size dry goods that populate our mundane existence have such dramatic effects on belief generating processes and embodied practices? And if they do, how exactly can the dualist semantics of the two-truths framework undo our ingrained hypostatization proclivities? The mereological nihilist singles out the effects on belief that configuration processes—by which various parts in the world come to be arranged chariot-wise and person-wise—can have when the demands of everyday life constrain one's ontological outlook. Speedy and effective long-distance travel may demand that one avails oneself of various modes of transportation. Should we admit cars, trains, and airplanes in our final ontology? Sure, the car in the commercial is more than a wheeled and motorized mode of transportation; it is also a symbol of comfort, elegance, efficiency, and status. It has its own event horizon, and it promises to augment human experience in profound ways. Nonetheless, such demands should be recognized for what they are: alluring invitations to reify that which is ultimately only a collection of parts.

The belief's acceptance may well reflect conventional linguistic practices that, once fixed, make possible assertions in which it is actually true that there are chariots and cars. The problem, as Siderits carefully explains, is that with such customary linguistic practices in place it becomes possible to assert both that there are such things as chariots and that there are certain parts that can be arranged chariot-wise. Semantic insulation between the two types of discourse may indeed give rise to anomalies, especially if one were to assert that both types of entities are real. But Siderits, following Thomasson (2007), thinks such anomalies indicate that the conditions under which something comes to be called a 'chariot' are not

always fixed by rules that apply to the discourse about parts, and so the existence of chariots, and of ordinary objects in general, can be vindicated.

However, it is an open question whether the dualist semantics of the two truths framework brings Buddhist Reductionism in line with the particular type that Parfit favors. Siderits thinks that it does, and considers a version of the teleportation scenario that aligns the two accounts, but in a way that does not eliminate the problem of indetermination. The question whether parts arranged a certain way are identical with or distinct from the chariot does not admit of a determinate answer, and neither does the question whether the swapping of parts from one to another chariot-wise arrangement results in an identical or different arrangement. Siderits' proposed solution involves metalinguistic analysis:

> What we can do is ascent to a meta-language and discuss those relations between our uses of chariot on the one hand and 'axle' and the like on the other that explain why, when the parts are assembled in a certain way, we say there is a chariot, and why there may be cases where we do not know whether to say there is a chariot, or to say that there is diachronic identity between some present chariot and an earlier chariot. We can talk about the whole, or we can talk about the parts, but we cannot talk about both in the same breath.
>
> *2015: 110*

The metalinguistic analysis solution may seem compelling enough in the case of chariots and other inanimate objects, but it confronts a paradox: since the parts taken together in some degree of combination do not add up to a chariot, there are no such things as chariots. Parfit confronts the paradox while answering an objection from Bernard Williams against the view that we can get personal identity from a psychological spectrum. The objection concerns the sorites paradox. Run in the reverse, the argument considers a scenario in which, by means of small enough surgical interventions, a surgeon could cause a person to cease to be psychologically continuous with previous stages of herself. If a surgeon could alter in small increments the neurochemistry of the brain in such a way as to profoundly transform the psychological features that make up who you are, does that mean you have ceased to exist? Parfit's response considers the semantics of vague terms, and his solution to the sorites paradox comes in the form of an appeal to stipulation: we simply decide how many grains of sand to call a heap. But whereas stipulation may work for natural kinds such as forests it does not work for artifacts such as chariots. A chariot is not simply an arbitrary number of parts, but rather a definitive set of parts arranged such as to function as a 'chariot'. These may vary depending on whether one considers extra features and embellishments, but one could agree on a minimum number of parts without which the concept chariot could not be meaningfully applied to parts arranged chariot-wise.

But even if we adopted such a provisory stance, does that mean the argument would work in the case of persons? If Parfit is right, the physical criterion is not

enough for personal identity, which is why we need to consider the full spectrum of psychological states. Since persons are defined primarily in terms of the subjective and phenomenal character of their conscious mental states, how does Buddhist Reductionism solve the personal identity problem? It would seem that any account of persons requires person-level descriptive categories, a point that, as we noted above, Parfit himself concedes.

> Even Reductionists do not deny that people exist. And, on our concept of a person, people are not thoughts and acts. They are thinkers and agents. I am not a series of experiences, but the person who has these experiences. A Reductionist can admit that, in this sense, a person is what has experiences or the subject of experiences. This is true because of the way we talk. What a Reductionist denies is that the subject of experiences is a separately existing entity distinct from a brain and body, and a series of physical and mental events.
>
> *1984: 223*

But in retaining a meaningful conception of 'persons' *as* thinkers and doers, Parfit's view departs from mainstream Buddhist Reductionism only to join the ranks of the Buddhist Personalists. As expounders of the 'personalist view' (*pudgalavāda*), Buddhist Personalists do agree with the general tenet of Buddhist Reductionism that there are no such things as enduring substantive selves. They also retain the bundle view and take 'persons' to be conceptualized in dependence upon the psychophysical aggregates. Where they differ, however, is in admitting that persons, although neither identical nor different from the aggregates, are nonetheless real. It is unclear whether for the Buddhist Personalists persons enjoy the same ontological status as the aggregates, in part because they exhibit novel properties that neither of the aggregates do. What is clear, however, is that although the Personalists admit that the notion of 'person' is conceived in dependence upon the aggregates, they do not regard it as a conceptual fiction. Sustained efforts to reject Buddhist Personalism as an orthodox position form an integral part of Buddhist Reductionism. Historically, Buddhist Personalists were understood to have forgone commitment to the no-self doctrine. However, it is now possible to accommodate their views on an emergentist or supervenience account of phenomenal properties.[11]

It is instructive that what makes Buddhist Personalism the target of criticism is a certain insistence on taking what we would now call the first-person perspective seriously, particularly as it finds articulation in discussions of the relation between consciousness and content. The key issue under dispute is the concept of person itself: if it stands simply for the collection of aggregates, then the Buddhist Reductionist is right to insist on its fictional status. However, if the concept of person is meant to capture something determinate, such as the view that persons are self-conscious subjects, then it is hard to see how the reductionist can make the case that the concept of 'person' lacks a proper referent. Consider, for

instance, Vasubandhu's refutation of personalism, and the argument that, since an individual can only possess 'person-properties' on account of possessing some sort of essence (*svabhāva*), and since Buddhist metaphysics is anti-essentialist, there is no ontological basis for admitting entities defined in terms of their possession of person-properties.[12] On the argument put forth here, person-properties are those very things that allow us to use the first-person pronoun meaningfully in sentences such as 'I am walking to my office' or 'I am thinking about the conversation we had yesterday'. The motivation for resisting such language use, according to Vasubandhu, reflects its unintended consequences: conceiving of the aggregates together as forming a person is a slippery slope to thinking there are such things as selves. To be clear, Vasubandhu does not reject the view that we do appear to ourselves to be more than just a body, or a bundle of feelings and thoughts. Rather, he rejects the notion that feelings, thoughts, and the body are something that we *possess*. Such a view would entail an existence apart from the aggregates. But since we are *nothing but* thoughts feelings, perceptions, memories, and bodies of a particular shape, size, and gender, no such residual existence is anywhere to be found.

In claiming that persons are conceived in *dependence upon* or *in reliance upon* the aggregates, though neither identical nor different from them, the Personalists invoke the *conceivability principle*. The dispute, then, hinges on whether or not the object of conception must be identical with its causal basis. If no such identity relation were to obtain, then conceiving would lack reference, and would not point to anything in the world. But even on a strongly non-referential view, conceiving must be regarded as an effective epistemic practice that entails, inter alia, reference to nonconceptual entities (regardless of their ontological status). On a referential view of conception, thus, to *conceive of* persons is to *conceive ourselves as* persons. In conceiving of becoming a father I conceived of myself as a father. In conceiving of unicorns, I conceive of them as possible or impossible. On the personalist view, the aggregates may serve as a causal basis for our conception of person without being identical with it. Conception is indeed reliant on the presence of some manifest phenomena, but conceiving of 'ourselves' in dependence upon the aggregates does not make 'us' identical with them. We conceive of fire in relation to fuel and of friendship in relation to common interest, but fire is not identical with fuel, nor is friendship identical with common interest. The Buddhist Personalist, pace Vasubandhu, wants to make the case that the object of conception is not always identical with its causal basis, which is why we can both retain a meaningful conception of person and be reductionists about selves.

It would seem that this particular version of Buddhist Reductionism, which retains the concept of person (though not that of self) would support Parfit's conception of personal identity in the STT case, though not in the ATT case. Since the person is conceived in dependence upon the causal series, though not identical with it, I am sufficiently connected to the previous instance of myself prior to entering the Earth STT. The conceivability criterion of personalism makes the ATT case problematic, given that there is no way to differentiate between conceiving of

myself in dependence upon the causal series now present on Mars as opposed to those still on Earth.
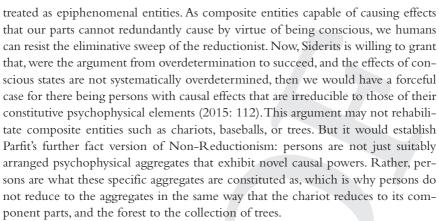
Might the person so conceived by Buddhist Personalists be a further fact, and could that be a mitigating factor against mereological nihilism? Siderits considers this objection by engaging some recent work in ontology, specifically Merricks' (2001) defense of eliminativism. Merricks' view is that strictly speaking there are no such things as statues, only "atoms arranged statuewise." In denying the existence of macroscopic objects, Merricks is not simply arguing against conventional uses of language, of the sort that take statement such as "there exists a statue" to mean "there exist atoms arranged statuewise" (2001: 3). Nor is he denying the existence of statues because statues are nothing over and above atoms arranged statuewise, and thus not mereologically distinct from their parts. Rather, he puts forward a complex argument from causal overdetermination: admitting that there exist such things as statues causally overdetermines their effects (because whatever effect can be attributed to a statue is already an effect of its microphysical parts). Since such overdetermination is not observed, we must admit that there are no such things as statues.[13] However, Merricks think the causal overdetermination argument does not apply in the case of human beings, because humans have causal powers that exceed those of their constitutive psychophysical parts. Most notably, humans instantiate *consciousness*, a property that does not supervene on the body's microphysical properties.

Merricks offers an interesting argument in support of this view operating on the premise that conscious and subjective mental properties bear the mark of being intrinsic. 'Intrinsic properties' are those properties "that an object can exemplify even if that object and its parts (if any) are the only objects that exist" (2001: 92). What is distinctive about humans is precisely that there is no metaphysical necessity that their intrinsic causal powers be implied by the existence of intrinsic properties of, and spatiotemporal and causal interaction among, their constituent elements. The argument goes as follow: suppose P, a conscious being, accidentally slices off her left index finger and thus shrinks. Suppose that at the very same moment of amputation, the atoms that compose her remain just as they were before amputation. The implication is that those atoms that are constitutive of P after amputation, are also constitutive of it immediately before amputation. Now, consider the pre-amputation finger-complement that is not identical with P (on the grounds that P has a part, the finger-complement, which the finger-complement itself simply lacks). If that is the case, and if we assume a relational view of consciousness (as a composite of atoms grouped by spatiotemporal and causal interrelations), then we must concede that before amputation there are two consciousnesses present, one for P and the other for the finger-compliment. But since we only ever have one such entity, P, the thesis that there are two consciousnesses must be false. Merricks thinks that to postulate the existence of a conscious pre-amputation finger-complement is to engage in an unacceptable multiplication of persons.

If it is true that the schema of the overdetermination argument does not apply to the effects a human causes by virtue of being conscious, then humans cannot be

treated as epiphenomenal entities. As composite entities capable of causing effects that our parts cannot redundantly cause by virtue of being conscious, we humans can resist the eliminative sweep of the reductionist. Now, Siderits is willing to grant that, were the argument from overdetermination to succeed, and the effects of conscious states are not systematically overdetermined, then we would have a forceful case for there being persons with causal effects that are irreducible to those of their constitutive psychophysical elements (2015: 112). This argument may not rehabilitate composite entities such as chariots, baseballs, or trees. But it would establish Parfit's further fact version of Non-Reductionism: persons are not just suitably arranged psychophysical aggregates that exhibit novel causal powers. Rather, persons are what these specific aggregates are constituted as, which is why persons do not reduce to the aggregates in the same way that the chariot reduces to its component parts, and the forest to the collection of trees.

The difference between Buddhist Reductionism and Buddhist Personalism should be clearer by now: whereas the reductivist claims that mental states strongly supervene on physical states and thus are reducible to the latter, the personalist takes them to possess novel causal powers that cannot be ontologically analyzed in terms of the properties of the physical states upon which they supervene. It is possible to interpret Buddhist Personalism as endorsing an emergentist view of persons and not simply a non-reductive physicalist view, but that would be dependent on how we interpret the 'in reliance upon' clause. Siderits' response to the arguments against causal overdetermination takes up the problem of compatibility between four different positions Buddhist Reductionism must consider when debating the ontological status of persons: (1) dualism, (2) reductive physicalism, (3) non-reductive physicalism, and (4) emergentism. When considered in the context of debates about wholes and parts, these positions correspond respectively to (1) mereological realism, (2) reductive mereological nihilism, (3) non-reductive mereological nihilism, and (4) emergentism (2015: 114). Drawing on contemporary debates on these issues,[14] Siderits admits that the non-reductive physicalist position (viz., mental states supervene on physical states but are not reducible to them) ought to, at least in principle, be compatible with emergentism (viz., mental states do supervene on physical states, but since they exhibit novel causal powers, physicalism must be false). The question, however, is whether conscious mental states, and by extension the person whose states they are, are fundamental despite being dependent upon the intrinsic properties of, and the spatiotemporal and causal interaction among, their constituent elements. Siderits thinks attempts to answer this question that appeal to the notion of downward causation face an explanatory gap, and so efforts to close that gap inevitably "push the non-reductivist in the direction of reductionism and identity" (2015: 116). So, the position is best classified as a version of the "non-reductive supervenience view" (2015: 117). But he does not agree that Buddhist Personalism offers a legitimate way to preserve the irreducibility of person-properties necessary for, say, negotiating moral desert, and finds functionalist accounts of mental content compelling enough the resist the non-reductionist move.

It is debatable whether the further-fact view of Buddhist Personalism is as inherently unstable as Siderits claims. In the next section, I will argue that it is not, and that considerations about self-consciousness threaten certain aspects of the Reductionist View, and thus render the conception of person (as conceived in dependence upon the aggregates, though neither identical nor different from them) essential for psychological accounts of personal identity.

Reflexivity, agency, and the unity of conscious experience

The much-quoted passage from *The Questions of Milinda* (*Milindapañha*) discussed above—in which Nāgasena argues against the ultimate reality of persons—stands in sharp contrast to an equally important, but less celebrated, critical statement from his interlocutor, King Milinda, concerning the moral and pragmatic consequences of the no-self doctrine:
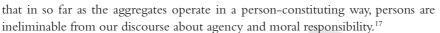
> If, most revered Nāgasena, no person is apprehended, then who gives you the offerings that you receive as a Buddhist monk? Who enjoys those things? Who practices disciplined conduct? Who enters into contemplation? Who attains to the goal of nirvāṇa, which is the outcome of the path of cultivation? If, revered Nāgasena, one was to slay you, his would not be the crime or murder.
>
> *Horner 1963: 100*

As this passage makes clear, the problem of personal identity is not exclusively metaphysical, but engages central issues in ethics and moral psychology. Do the moral and mental forms of cultivation at the heart of Buddhist practice demand a robust notion of agency? Buddhism is indeed unique in articulating a theory of action that, it seems, dispenses altogether with the notion of agent causation. Even though practices of moral and mental cultivation form an integral part of the Buddhist path, there is no stable self or agent who bears the accumulated responsibility for initiating those pursuits, and seemingly no normative framework against which some dispositions, thoughts, and actions are deemed felicitous, and thus worthy of cultivation, while others are not so deemed. The agent-neutral metaphysical picture of Buddhist scholasticism thus challenges traditional conceptions of moral agency. One way to address this challenge is to claim that the discourse of 'persons' and the discourse of 'causes' are compatible in so far as they belong in two distinct and incommensurable domains.[15] Compatibilists who adopt this position typically cite evidence from social and cognitive psychology to show that any robust notion of agent causation must be incoherent.[16] But, as I have argued elsewhere, such solutions compromise traditional conceptions of moral responsibility and render ethical conduct indistinguishable from merely pragmatic acts (Coseru 2016). Indeed, despite the dominance of an ultraminimalist account of agency, there are good and compelling reasons to give Buddhist Personalists credit for insisting

that in so far as the aggregates operate in a person–constituting way, persons are ineliminable from our discourse about agency and moral responsibility.[17]

Does Buddhist Personalism provide a closer analogue for Parfit's theory of personal identity (and its psychological reductionism) than Buddhist Reductionism, with its stated mereological nihilist view that there are no such things as composite entities? I think that it does, for conceivability reasons: the psychological criterion is plausible only to the extent that I can conceive of myself in dependence upon a sufficiently similar series of aggregates. But the conceivability principle concerns the epistemological, rather than the ontological, dimension of personal identity, which brings up an altogether different set of considerations, specifically about the relations that obtain between self-referential mental states (those that presuppose the notion of oneself as an agent) and self-consciousness. The epistemological dimension is framed by a different set of questions that pertain not to what awareness supervenes on but to its structure and specific properties, specifically: What in particular accounts for a mental state becoming an instance of self-consciousness? Does self-consciousness require that its self-referential relation is present to itself as an object? If we can answer these questions, we can make progress in understanding the relation between self-referential mental states and self-consciousness. And if we can get clarity about the nature and character of self-consciousness, we are in a better position to understand what it is that makes us persons.[18]

Recall the schematic analysis of the five aggregates that informs the Buddhist Reductionist account of personal identity. In this analysis, only the body is a physical aggregate *stricto sensu*. Feelings, perceptions, dispositions, and consciousness *can* acquire an objective aspect, but are not properly speaking empirically tractable phenomena. Nor are they things, that is, abstract entities with well-defined properties and functional characteristics. What this reductive analysis of persons in terms of their constitutive features is meant to capture is not what persons are *made of*, but rather what human experience is *constituted as*: specifically, as a series of intentional and self-referential mental events. Consider the paradigmatic example of pain: as a sensation, pain is not reducible to the physical substrate, say a finger, in which it is instantiated (nor presumably to a mere physiological response). Rather, pain is constituted as a distinctly qualitative phenomenon whose intentional content cannot be dissociated from its subjective aspect. There is no such thing as generic or impersonal pain (understood strictly in terms of, say, the activation of Aδ- and C- fibers following an intense stimulation of nociceptors) apart from phenomenologically foregrounded sensations of some kind: of burning, stinging, or throbbing. Feelings may define the quality of the impressions that result from contact with an object, with the implication that they perhaps stand in a causal relation with these objects. In the schematic analysis of Buddhist scholasticism, they are categorized as mental states conditioned by habitual tendencies (*vāsanā*), which, in turn, they condition. Likewise, apperception (*saṃjñā*), the capacity to make intelligible or cause to be understood, although dependent on a multiplicity of psychological factors, captures the datum of experience only as unified into a single percept (since what Abhidharma psychology understands as a
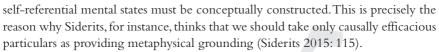
simple apprehension and identification of a sensible trope actually involves a significant amount of unification of sensory input that goes on behind the scenes). Volitions too fit the same profile, with one important difference: rather than attending to the object at hand or providing a sort of transcendental unity of apperception, they bring forth future states of existence. As dispositions to act in certain ways, they cleave the mental domain into two classes of conditioned phenomena: those that are internal to consciousness, such as, for instance, obsessive dispositions like greed and delusion, and those that are dissociated from it, usually taken to refer to latent dispositions typically comprising various biological and physical traits.[19]

This dynamic model of personal identity is not incompatible with the notion that there are phenomenally primitive features of experience, features that any analysis of the structure of self-consciousness must ultimately reveal. When the Buddhist Personalists insist that a mere functional account of the aggregates will not suffice to explain why an action counts as murder they draw attention to the specificity and individuality of a given bundle of aggregates, and hence, of its actions and consequences: "If the person were absolutely non-existent, then there could not be killing nor (nor?) would the killer have killed anything. There would be nothing like theft and robbery … good and bad would yield neither freedom nor bondage; even bondage would have no one bound. There would be neither the doer nor the deed not any result thereof" (Venkataramanan 1953: 177). Indeed, understanding why something is categorized as murder and not simply as the rearranging of aggregates presupposes a conception of intentional action that is unintelligible without reference to persons. If living beings were mere conventional designations, as Nāgasena contends, then there would be no non-arbitrary way to assign guilt in the case of killing a sentient being, since killing a living ox would be no different from destroying a clay ox.[20] Unlike clumps of clay, living beings are characterized primarily in terms of their capacity for responsive and intentional action: they can both do things and have things done to them in a way unavailable to insentient objects. Persons, unlike mountains or chariots, are not simply generic unities of aggregates, but self-disclosing wholes. Persons are what they are by virtue of the fact that their aggregates are perceived to belong together.[21] Aggregates grouped together in a person-constituting way are specific to themselves in a way that mountains grouped together or chariot parts assembled together are not.

What are the implications of this position for Buddhist Reductionism in particular, and for the Reductionist View of personal identity in general? There are various responses to this question. One may take the reflexive structure of self-consciousness to point to deeper aspects of consciousness as conscience,[22] or delineate several ways in which pronominal and indexical uses of the 'I' can engender a false sense of self.[23] One may also claim that the perceived unity of consciousness results from certain patterns of attention to, and grasping of, various features of experience.[24] Whether one takes self-consciousness to have a distinctive structure and clearly specifiable content, for the Buddhist Reductionist no minimal account of agency can escape the antirealist stance of the no-self doctrine and its implications, namely that all modes of 'I-making' (and 'I-sensing') characteristic of

self-referential mental states must be conceptually constructed. This is precisely the reason why Siderits, for instance, thinks that we should take only causally efficacious particulars as providing metaphysical grounding (Siderits 2015: 115).

Buddhist Reductionism does indeed pose a significant challenge for conceptions of phenomenal consciousness that take consciousness to consist in more than just a succession of associated ideas or a construct of post-hoc rationalization. However, by treating its contents as transient episodes arising within a continuum of caus-ally interconnected states, Buddhist Reductionism itself confronts an epistemic explanatory gap: if all there is to the experiential domain is a stream of momentary, object-directed mental events, just what it means for an object to be known, and by whom? Can Buddhist Reductionism, with its account of primitive atom-like 'qualitons' of experience, provide an adequate conceptual basis for mapping out the distinctly self-intimating character of knowledge episodes? To answer this question let me briefly consider an epistemological attempt to negotiate the difference between egological and non-egological accounts of self-consciousness, drawing principally on the work of Dignāga.

Building on the Abhidharma analytic project, Dignāga, not unlike Brentano, advances a dual-aspect theory of mind, which takes the subjective or qualitative aspect (*svābhāsa*) and the object-oriented or intentional aspect (*viṣayābhasa*) to be constitutive features of the structure of cognitive awareness.[25] In large measure, this theory rests on three distinct claims: (1) that we are directly aware of each occurrent mental state; (2) that each occurrent mental state has a dual aspect: it has both sub-jective and objective content; and (3) that each occurrent mental state is also reflex-ively self-conscious. The first claim goes against the view that occurrent mental states are ultimately impersonal or anonymous. To hold such a view would be akin to claiming that experience lacks a distinctly subjective character or possesses such character only by virtue of distortions embedded in the structure of 'I-making' and 'self-grasping' tendencies.[26] The second claim identifies subjectivity and inten-tionality as distinct structural features rather than discrete elements of the mental domain. The subjective aspect is constitutive of an implicit openness to what is given, while the objective aspect captures what the mental state is about: an object or mental content of some kind. Finally, the third claim is intended to capture the mode of presentation of all conscious cognitive states. Incidentally, the first and third claims can also be read as making the case that effortful self-knowledge—of the sort gained through introspection, intersubjective reports, or reflective attitudes—depends on tacit or non-propositional modes of acquaintance.

This understanding of the structure of conscious experience is not unlike that put forward by those who argue that our subjectivity is immersive rather than egological. Consider, for instance, this statement from Zahavi, who takes Heidegger's lead on this issue: "I am acquainted with myself when I am captured and captivated by the world. Self-acquaintance is indeed only to be found in our immersion in the world, that is, self-acquaintance is always the self-acquaintance of a world-immersed self" (Zahavi 2005: 82). On this view, self-acquaintance is not some-thing that occurs apart from our immersion in the world and its complex set of

intentional and intersubjective relations. Reflecting on the mode of presentation of our mental states, P. F. Strawson likewise observes, "our desires and preferences are not, in general, something we just note in ourselves as alien presences. To a large extent they are we" (Strawson 1992: 134). To claim, thus, that access to our mental lives is always mediated in some fashion or another, perhaps by participation in a shared domain of language and reflection—is to ignore these essential features of self-consciousness.
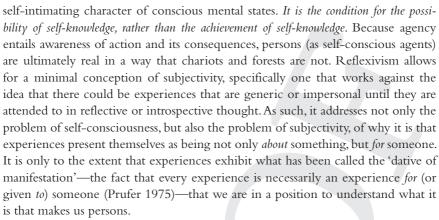
Dignāga's account of reflexive self-consciousness (*svasaṃvitti*) does raise some legitimate concerns, for instance, about the absurd consequence of admitting that all cognitions, by virtue of being self-presenting, are epistemically warranted. Yet the main issue, as I have argued at length elsewhere (Coseru 2012, Chapter 8), is specifying the criteria under which occurrent mental states become epistemically salient. A reductive causal account that treats consciousness as emerging from elements that themselves lack phenomenal properties is necessarily incomplete if it does not explain how those elements acquire the phenomenal properties that they do when grouped together. Of course, not all groupings of microphysical elements need necessarily exhibit a reflexive structure.[27] But those that do call for an explanation that considers the whole person and not just its constitutive elements. In putting forward an account of consciousness as self-intimating, Dignāga is thus concerned to explain how a conscious mental episode, which is irreducible, can become the vehicle of self-knowledge. In that sense, he shares a common ground with, inter alia, Aristotle, Descartes, Locke, Brentano, Husserl, Sartre, and many contemporary defenders of the view that the nature of consciousness is such that all conscious mental states are non-representationally or pre-reflectively aware of themselves.[28] The picture they present corresponds to what some have termed non-reductive "one-level theories of consciousness": that is, theories which propose that consciousness is essentially a matter of having or being an awareness of a world that does not require a prior (representational) awareness of our own mental states (Thomasson 2008).

Reflexivism or the thesis that self-consciousness consists in mental states being implicitly conscious of their occurrence thus serves as a grounding principle, enabling the intentional and subjective aspects of experience to emerge co-constitutively in each instance of cognitive awareness. Can such reflexive self-consciousness provide enough metaphysical grounding for thoughts of the sort 'I am in pain'? Do such self-ascriptions require independent criteria (or the normative, indexical, or ownership kind) for individuating streams of conscious episodes to manifest as occurring *for me*? Can conscious mental states occur without any sense of whose states they are? Can there be consciousness without self-consciousness?

If reflexive self-consciousness has a distinctive character, a specific givenness or for-me-ness, then even misascriptions (as in the case of thought insertion or various pathologies of the self such as ego dissolution) would be unintelligible if thought were transparent with regard to its occurrent for-me-ness (Bermudez 1998; Zahavi & Kriegel 2015). Reflexivism, then, is simply a statement about the

self-intimating character of conscious mental states. *It is the condition for the possibility of self-knowledge, rather than the achievement of self-knowledge.* Because agency entails awareness of action and its consequences, persons (as self-conscious agents) are ultimately real in a way that chariots and forests are not. Reflexivism allows for a minimal conception of subjectivity, specifically one that works against the idea that there could be experiences that are generic or impersonal until they are attended to in reflective or introspective thought. As such, it addresses not only the problem of self-consciousness, but also the problem of subjectivity, of why it is that experiences present themselves as being not only *about* something, but *for* someone. It is only to the extent that experiences exhibit what has been called the 'dative of manifestation'—the fact that every experience is necessarily an experience *for* (or given *to*) someone (Prufer 1975)—that we are in a position to understand what it is that makes us persons.

We may conclude that what makes the further fact view of there being persons plausible is the structure of consciousness itself. This way of understanding persons is not unlike Locke's view of persons as self-conscious rational subjects. Neither a biological entity, nor an immaterial soul, a person is what it is by virtue of the fact that it,

> can consider it self as it self, the same thinking thing, in different times and places; which it does only by that consciousness, which is inseparable from thinking, and, as it seems to me essential to it: It being impossible for any one to perceive, without perceiving, that he does perceive. When we see, hear, smell, taste, feel, meditate, or will any thing, we know that we do so. … And as far as this consciousness can be extended backwards to any past Action or Thought, so far reaches the identity of that Person.
>
> *Locke 1689 2.27.9, 335*

The Buddhist Personalist may eschew the requirement that the self-conscious subjectivity in play be manifest only in thought as rational deliberation and action, but agree that without self-consciousness it is hard to understand how self-concern, which is essential for moral cultivation, would be possible. For in looking to reward or punish persons for their deeds we do not look for a particular body, mental state or an immaterial substance of some sort, but rather "to connections between the consciousness of the one who did the deed and the one who is being held accountable." (Ainslie and Ware 2014: 248).[29] Buddhist Personalists may also agree with Hume and the Parfitian version of the Reductionist View that our perceptions or sensible intuitions reveal no such thing as a fixed or stable self. Awareness of the inner contents of our mind lacks a necessary relation to a unified self. But while it may be true that we do not need to be self-conscious in a higher order or metacognitive sense in order to perceive and think, the operations of perception and thought do presuppose that conscious experiences are unified in such a way that their contents can be synthesized. And it is the unity of conscious experience, achieved in the constitutive operations of the aggregates, that Buddhist Personalists point to when they claim that persons are ultimately real.

Conclusion

Parfit's motivation for rejecting the Non-Reductionist View concerns the claim that there is more to being a person than what accounts of physical and psychological continuity can satisfactorily provide. The Non-Reductionist makes the ultimately unsupported claim that "the unity of consciousness at any time is explained by the fact that several experiences are being had by a person" (1984: 275). Split brain cases and teleportation thought experiments, however, make it hard to explain unity by insisting on the importance of maintaining personal identity over time across different streams of consciousness. In conceiving of unity in terms of the ascription of different experiences to a particular subject, we are confronted with the unwarranted assumption that "there can be, in a person's life, subjects of experiences that are not persons" (ibid.). Such cases are better explained, argues Parfit, by appealing to the Reductionist Psychology Criterion, which claims that "at any time, there is one state of awareness of the experiences in one stream of consciousness, and another state of awareness of the experiences in the other stream" (ibid.).

In developing the criteria for psychological reductionism, Parfit's theoretical perspective points in the direction of a formal but variant structure of awareness. This structure ought to, at least in principle, be able to accommodate continuity without the sameness or identity that ascribing such states to something like a Cartesian Ego would entail. It is unlikely that Buddha and most Buddhist Reductionists would have agreed with this part of his theory. Although it is in keeping with the principles of the Buddhist no-self view, Parfit's insistence on an external criterion for individuating or indexing awareness to a particular stream of consciousness may be at odds with what I have argued are certain salient and ineliminable features of phenomenal consciousness, specifically its self-reflexivity. If self-reflexive conscious episodes lack an intentional structure, then they cannot provide the minimal sense of internal distance necessary for subjectivity. On the dual aspect model of consciousness sketched above (see also Coseru 2015, 66, 79), mental streams are differentiated by being covariant with intentional behavior, which presupposes that subjectivity and intentionality are structural features of consciousness, rather than constructs or relations among discrete mind moments (as presupposed by mereological nihilists).

This chapter has considered whether Buddhist Reductionism provides support for Parfit's psychological criterion for personal identity given considerations about the seemingly irreducible character of phenomenal consciousness. Taking Buddhist Personalism as a point of reference, I have argued that reductionism about selves can be reconciled with the seemingly irreducible character of self-consciousness. Furthermore, Buddhist philosophers who follow in the footsteps of Dignāga and his successors provide a viable, if not altogether unproblematic, model for how this reconciliation might proceed. Whether or not this epistemological project is constrained by ontological assumptions about the constitutive elements of persons is open to debate. What is less controversial is that in rejecting a permanent locus for experience, the Buddha created an opportunity for the problem of personal identity

to be explored not only on metaphysical but on empirical and phenomenological grounds as well. What sets apart non-reductive theories of persons as self-conscious subjects from their alternatives, then, is the notion that personal identity is not really a matter of the persistence or continuity of enough of the constitutive (physical and psychological) elements of persons, as demanded by the psychological criterion. Rather, on the model put forward here, it is a matter of understanding that the principle of individuation is neither external nor transcendent but internal to the structure of consciousness and its operations.

Notes

1 Within philosophy of mind, reductionism is concerned mainly with solving the mind–body problem, rather than, as was the case with Carnap (1934) and Neurath (1931), with the unity of scientific theories. Hence, the main motivation for reductionism with regard to questions of personal identity is ontological parsimony.

2 In the case of intertheoretic reduction, the question is whether the atomic and continuum scale models can be bridged. While some reductionists think that in principle it is possible to derive continuum scale entities from atomic scale details (Batterman 1995), others (in particular those who favor top-down modeling) take the view that what we can hope for at most is identifying a set of intertheoretic relations (Berry 2001) rather than a hierarchy of theories operating at different scales. The upshot of this debate is that philosophical theories of reduction, which are typically modeled on scientific theories, must come to terms with the fact that reduction is problematic even in the case of the natural sciences.

3 One way to understand the difference between the non-reductionist and the reductionist views of personal identity is along the simple/complex divide: the non-reductionist favors the simple, soul or Cartesian Ego, view, whereas the reductionist prefers the complex view that entails relations among physical and psychological states. Holding a soul view, of course, does not necessarily amount to holding a brute fact view, although in the absence of non-circular criteria for personal identity (of the sort required by the complex view) it is hard to tell them apart. What motivates recent defenders of the simple view (e.g., Baker 2013; Lowe 2009, 2013; Nida-Rümelin 2013, Swinburne 2013) is not commitment to a Cartesian Ego, but rather the notion that a specific, perhaps non-conceptual and pre-reflective, type of self-awareness seems indispensable to framing any account of personal identity. I address this point below, in §4.

4 Parfit supports his claim regarding indeterminacy with the case of the Combined Spectrum, which, as the name suggests, involves all the possible variations of physical and psychological connectedness between such substantially altered person stages that the person I am now would in no way be connected with the one at the further end of the spectrum.

5 The first to draw attention to their similarity was Collins (1985), though in a subsequent article (Collins 1997) he softened his position, regarding the two views as similar rather than identical. Similar attempts to draw together the two views and to adopt a Parfitian conceptual schema in pursuing various projects in Buddhist philosophy are found in Bastow (1986), Stone (1988, 2005), Giles (1993), Duerlinger (1993), Basu (1997), Siderits (1997, 2003), Perrett (2002), Ganeri (2007), and Sauchelli (2016).

6 Parfit first articulated his views on personal identity as early as his 1971 paper "Personal Identity," so that by the time *Reasons and Persons* was published, his views were widely known, and had already generated a great deal of discussion.
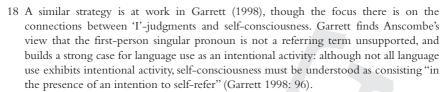
7 Parfit gets around the problematic issue of the relation between episodic and semantic memory by appealing to *q*-memory (quasi memory—a new category of memory first introduced by Shoemaker (1970)). According to Parfit, "I am q-remembering an experience if (1) I have a belief about a past experience which seems in itself like a memory belief, (2) someone did have such an experience, and (3) my belief is dependent upon this experience in the same way in which a memory of an experience is dependent upon it" (1971: 15).

8 This text, along with Vasubandhu's *Treasury of Metaphysics* (*Abhidharmakośabhāṣya*) and Buddhaghosa's *Path to Purity* (*Visuddhimagga*), serve as his main source of Buddhist ideas about personal identity. Parfit includes several passages from these texts, excerpted from Stcherbatsky's (1919) seminal work on the Buddhist no-self doctrine. He also cites Collins' (1982) landmark study of the early Buddhist doctrine of *anatta*. See also Collins (1997) for a response that considers the extent to which support for a reductionist account of personal identity on the Parfitian model can indeed be found in the Buddhist tradition.

9 For an overview of the impact that Parfit's categorical framework and theoretical perspectives (on metaphysics and ethics) have had on Buddhist philosophy, see Hanner (2018).

10 All references are to the second edition, Siderits (2015).

11 Buddhist Personalists (Pudgalavāda) include the Vātsīputriyas and the Sāṃmitīyas, though their exact views on persons are hard to reconstruct since they are mostly preserved in unsympathetic treatments by their critics, mainly in the *Points of Controversy* (*Kathāvatthu*), Vasubandhu's *Treasury of Metaphysics* (*Abhidharmakośabhāṣya*), Harivarman's *True Attainment Treatise* (*Satyasiddhiśāstra*), Devasarman's *Body of Consciousness* (*Vijńānakāya*), and Śāntarakṣita's *Compendium of Principles* (*Tattvasaṃgraha*). The only source that presents their views unbiasedly is the *Sāṃmitiya Nikāya Śāstra*, now preserved only in Chinese (an English translation in Venkataramanan (1953)). The prevalent interpretation, gleaned from the *Points of Controversy*, is that Pudgalavādins take persons to be ultimately real. According to most of the other sources, Pudgalavādins think persons occupy an ontological position somewhere between conventional and ultimate reality. See Priestly (1999) for detailed studies of Buddhist Personalism, and also Châu (1999) for a general survey and reconstruction of their views.

12 The refutation, known as *Treatise on the Negation of the Person*, is appended as Chapter 9 of Vasubandhu's *Treasury of Metaphysics* (*Abhidharmakośabhāṣya*). See Duerlinger (2003) for a detailed treatment of this section of Vasubandhu's treatise.

13 For a response to the argument from overdetermination, see Sider (2003).

14 In particular, Crane (2010), Noordhof (2010), and Kistler (2010).

15 Proponents of this move include, inter alia, Siderits (1997, 2008), Flanagan (2002), and Meyers (2014).

16 See Caruso (2012), Smart (2006), and Wegner (2002) for various attempts to prove the illusory nature of experiences of mental causation. While not conclusive, Nahmias et al. (2004) review experimental data that seems to favor compatibilist over incompatibilist accounts of free will.

17 See Carpenter (2015) for a sympathetic treatment of Buddhist Personalism that proposes a new, constitutive or developmental model for understanding the Pudgalavāda, aimed at extending the categorical framework of Buddhist metaphysics. On this model, the aggregates (of perception, consciousness, etc.) relate to each other in a person-constituting way, which she regards as "the most minimal explanation possible" that must be accepted if we are "to do justice to the phenomena *without* adopting a substantialist theory of self" (2015: 27).

18 A similar strategy is at work in Garrett (1998), though the focus there is on the connections between 'I'-judgments and self-consciousness. Garrett finds Anscombe's view that the first-person singular pronoun is not a referring term unsupported, and builds a strong case for language use as an intentional activity: although not all language use exhibits intentional activity, self-consciousness must be understood as consisting "in the presence of an intention to self-refer" (Garrett 1998: 96).

19 See, e.g., *Treasury of Metaphysics* (*Abhidharmakośabhāṣya*) 2.23–34 in de la Vallée Poussin (1980: 45–62).

20 For Harivarnan's discussion of some of the ethical implications of the no-self view in *Satyasiddhiśāstra*, see Priestly (1999: 96).

21 Carpenter (2015: 23f) thinks the "belonging together" clause is not just a matter of spatiotemporal proximity, a feature shared by all aggregates that belong together, but something having to do with their functionality and development, even though in the end she admits that functionality alone is not sufficient to distinguish persons from chariots.

22 This is the strategy adopted by Harvey (1995).

23 Collins (1982) offers precisely such an analysis.

24 This position, first articulated by Albahari (2006), is developed at length by Ganeri (2017).

25 See, e.g., *Compendium of Epistemology* (*Pramāṇasamuccaya*) 1.10, in Hattori (1968: 29). See also Coseru (*forthcoming*) for a discussion of Dignāga's theory relative to the problem of self-knowledge.

26 This view is attributed to Asaṅga. As he observes in the *Compendium of the Great Vehicle*:

> How does one know that mind (*manas*) in the sense of 'afflicted mind' (*kliṣṭa-manas*) exists? Without it, there would be no uncompounded ignorance, that is, a basic ignorance not yet associated with all the faults (*doṣa*), but serving as their base (*āśraya*). Indeed, introspective awareness (*mano-vijñāna*) must also have a simultaneous basis, as do the five types of empirical consciousness whose support are their material organs. Such a simultaneous support can only be found in the 'afflicted mind.'
>
> *Mahāyāna-saṃgraha 1.6–7 in Lamotte 1938: 36*

27 Micropsychism, roughly the view that human consciousness derives from a combination of micro-conscious entities, might be taken as the exception here (see, e.g., Strawson 2006). But micropsychism faces the well-known combination problem (Seager 1995): how exactly do micro-level entities with their own very basic forms of consciousness combine together to form human consciousness? In response to the combination problem, some defenders of panpsychism reject the atomistic metaphysics of micropsychism in favor of cosmopsychism, which holds that individual human consciousness is derived from, or an expression of, a single unified field of consciousness (Jaskolla & Buck 2012; Shani 2015, 2018; Nagasawa & Wager 2016; Goff 2017). Cosmopsychism faces the opposite problem, namely the decombination problem: how do we get from a unified field of consciousness to individual human and animal consciousness.

28 See Strawson (2015) for a defense of the self-intimation thesis that draws together these major figures.

29 As Ainslie and Ware make quite clear, Locke not only defends the view, typically associated with Sartre, that consciousness is reflexive (that is, consciousness entails self-consciousness), but also (and more controversially) that consciousness is not "limited simply to our occurrent mental states" but rather extends its contents from one conscious moment to the next (2014: 249).
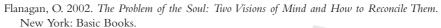
References

Ainslie, D. C. and Ware, O. 2014. "Consciousness and personal identity". In A. Garrett (ed.), *The Routledge Companion to Eighteenth Century Philosophy*. Routledge, 245–264.

Albahari, M. 2006. *Analytical Buddhism: The Two-Tiered Illusion of Self*. Hampshire: Palgrave Macmillan.

Baker, L. R. 2013. "Personal identity: a not-so-simple view". In G. Gasser and M. Stefan, *Personal Identity: Complex or Simple*. Cambridge: Cambridge University Press, 179–191.

Bastow, D. 1986. "Self-construction in Buddhism". *Ratio* 28(2): 97–113.

Basu, A. 1997. "Reducing concern with self: Parfit and the ancient Buddhist schools". In D. Allen (ed.), *Culture and Self: Philosophical and Religious Perspectives, East and West*, 97–109. Boulder, CO: Westview Press.

Batterman, R. W. 1995. "Theories between theories: Asymptotic limiting intertheoretic relations". *Synthese* 103: 171–201.

Bermudez, J. L. 1998. *The Paradox of Self-Consciousness*. Cambridge, MA: MIT Press.

Berry, M. W. 2001. "Chaos and the semiclassical limit of quantum mechanics (is the moon there when somebody looks?)". In *Quantum Mechanics: Scientific Perspectives on Divine Action*, R. J. Russell, P. Clayton, K. Wegter-McNelly and J. Polkinghorne (eds.), Vatican Observatory CTNS Publications, 41–54.

Carnap, R. 1934. *The Unity of Science*. London: Kegan Paul, Trench, Trubner, and Co.

Carpenter, A. 2015. "Persons keeping their karma together: the reasons for the Pudgalavāda in early Buddhism". In K. Tanaka, Y. Deguchi, J. Garfield and G. Priest (eds.) *The Moon Points Back*. New York: Oxford University Press, 1–44.

Caruso, G. D. 2012. *Free Will and Consciousness: A Determinist Account of the Illusion of Free Will*. Lanham: Lexington Books.

Cassam, Q. 1995. "Introspection and bodily self-ascription". In J. L. Bermudez, A. Marcel and N. Eilan (eds.) *The Body and the Self*. Cambridge, MA: MIT Press.

Châu, T. T. 1999. *The Literature of the Personalists of Early Buddhism*. Translated by S. Boin-Webb. Delhi: Motilal Banarsidass.

Collins, S. 1982. *Selfless Persons: Imagery and Thought in Theravāda Buddhism*. Cambridge: Cambridge University Press.

Collins, S. 1997. "A Buddhist debate about the self; and remarks on Buddhism in the work of Derek Parfit and Galen Strawson". *Journal of Indian Philosophy* 25(5): 467–493.

Coseru, C. 2012. *Perceiving Reality: Consciousness, Intentionality, and Cognition in Buddhist Philosophy*. New York: Oxford University Press.

Coseru, C. 2015. "Perception, causally efficacious particulars, and the range of phenomenal consciousness". *Journal of Consciousness Studies* 22(9–10): 55–82.

Coseru, C. 2016. "Freedom from responsibility: agent-neutral consequentialism and the Bodhisattva ideal". In R. Repetti *Buddhist Perspectives on Free Will: Agentless Agency?* New York: Routledge, 92–105.

Coseru, C. (forthcoming). "Whose consciousness? reflexivity and the problem of self-knowledge". In M. Siderits, K. Ching and J. Spackman *Buddhist Philosophy of Consciousness*.

Crane, T. 2010. "Cosmic hermeneutics vs emergence: the challenge of the explanatory gap". In G. Macdonald and C. Macdonald (eds.), *Emergence in Mind*. Oxford: Oxford University Press, 22–34.

de la Vallee Poussin, L. 1980. *L'Abhidharmakośa de Vasubandhu*. Vol. 1. Bruxelles: Institut Belge des Hautes Études Chinoises.

Duerlinger, J. 1993. "Reductionist and nonreductionist theories of persons in Indian Buddhist philosophy". *Journal of Indian Philosophy* 21(1): 79–101.

Duerlinger, J. 2003. *Indian Buddhist Theories of Persons: Vasuabndhu's "Refutation of the Theory of a Self"*. London: Routledge Curzon.

Flanagan, O. 2002. *The Problem of the Soul: Two Visions of Mind and How to Reconcile Them*. New York: Basic Books.

Ganeri, J. 2007. *The Concealed Art of the Soul: Theories of Self and Practices of Truth in Indian Ethics and Epistemology*. Oxford: Oxford University Press.

Ganeri, J. 2017. *Attention, Not Self*. Oxford: Oxford University Press.

Garrett, B. 1998. *Personal Identity and Self-Consciousness*. London: Routledge.

Giles, J. 1993. "The no-self theory: Hume, Buddhism, and personal identity". *Philosophy East and West*, 43(2): 175–200.

Goff, P. 2017. *Consciousness and Fundamental Reality*. Oxford: Oxford University Press.

Hanner, O. 2018. "Buddhism as reductionism: personal identity and ethics in Parfitian readings of Buddhist philosophy; from Steven Collins to the present". *Sophia* 57: 211–231.

Harvey, P. 1995. *The Selfless Mind: Personality, Consciousness and Nirvāṇa in Early Buddhism*. Richmond: Curzon Press.

Hattori, M. 1968. *Dignāga, On Perception*. Cambridge, MA: Harvard University Press.

Horner, I. B. 1963. *Milinda's Questions*. Vol. I. London: Luzac & Company.

Jaskolla, L. J. and Buck, A. J. 2012. "Does panexperientialism solve the combination problem". *Journal of Consciousness Studies*, 19(9–10): 190–199.

Kirchin, S. (ed.) 2017. *Reading Parfit*. London: Routledge.

Kistler, M. 2010. "Strong emergence and freedom—comment on A. Stephan". In G. Macdonald and C. Macdonald (eds.), *Emergence in Mind*. Oxford: Oxford University Press, 240–251.

Lamotte, E. 1938. *La Somme du Grand Véhicule d'Asaṅga (Mahāyānasaṃgraha)*. Louvain-la-Neuve: Université de Louvain.

Locke, J. 1689/1975. *An Essay Concerning Human Understanding*. P. Nidditch (ed.). Oxford: Oxford University Press.

Lowe, E. J. 2013. "The probable simplicity of personal identity". In G. Gasser and M. Stefan, *Personal Identity: Complex or Simple*. Cambridge: Cambridge University Press, 137–155.

Merricks, T. 2001. *Objects and Persons*. Oxford: Clarendon Press.

Meyers, K. 2014. "Free persons, empty selves: freedom and agency in light of the two truths". In M. Dasti and E. F. Bryant (eds.) *Free Will, Agency, and Selfhood in Indian Philosophy*. New York: Oxford University Press, 41–67.

Nagasawa, Y. and Wager, K. 2016. "Panpsychism and priority cosmopsychism". In P. Brüntrup and R. Jaskolla (eds.), *Panpsychism*. New York: Oxford University Press, 113–129.

Nahmias, E., Morris, S., Nadelhoffer, T. and Turner, J. 2004. "The phenomenology of free will". *Journal of Consciousness Studies*, 11(7–8): 162–179.

Neurath, O. 1932/1983. "Sociology in the framework of physicalism'. In Neurath, M. *Philosophical Papers 1913–1946*, R. S. Cohen and M. Neurath (eds.), Dordrecht: Reidel, 58–90.

Nida-Rümelin, M. 2013. "The non-descriptive individual nature of conscious beings". In G. Gasser and M. Stefan, *Personal Identity: Complex or Simple*. Cambridge: Cambridge University Press, 157–176.

Noordhof, P. 2010. "Emergent causation and property causation". In G. Macdonald and C. Macdonald (eds.), *Emergence in Mind*. Oxford: Oxford University Press, 69–98.

Parfit, D. 1971. "Personal identity". *The Philosophical Review*, 80(1): 3–27.

Parfit, D. 1984/87. *Reasons and Persons*. Oxford: Clarendon Press.

Parfit, D. 1999. "Experiences, subjects, and conceptual schemes". *Philosophical Topics* 26(1–2): 217–270.

Parfit, D. 2002. "Reductionism and personal identity". In D. J. Chalmers (ed.), *Philosophy of Mind: Contemporary Readings*. Oxford: Oxford University Press, 655–661.

Perrett, R. 2002. "Personal identity, minimalism, and Madhyamaka". *Philosophy East and West* 52(3): 373–385.

Priestley, L. 1999. *Pudgalavāda Buddhism: The Reality of the Indeterminate Self*. Toronto: Centre for South Asian Studies, University of Toronto.

Prufer, T. 1975. "An outline of some Husserlian distinctions and strategies, especially in the Crisis". *Phänomenologische Forschungen* 1: 189–204.

Sauchelli, A. 2016. "Buddhist reductionism, fictionalism about the self, and Buddhist fictionalism". *Philosophy East and West* 66(4): 1273–1291.

Shani, I. 2015. "Cosmopsychism: a holistic approach to the metaphysics of experience". *Philosophical Papers* 44(3): 389–417.

Shani, I. 2018. "Beyond combination: how cosmic consciousness grounds ordinary experience". *Journal of the American Philosophical Association*, 16(41): 390–410.

Shoemaker, S. 1970. "Persons and their pasts". *American Philosophical Quarterly* 7(4): 269–285.

Sider, T. 2003. "What's so bad about overdetermination?" *Philosophy and Phenomenological Research* 67(3): 719–726.

Siderits, M. 1987. "Beyond compatibilism: a Buddhist approach to freedom and determinism". *American Philosophical Quarterly* 24(2): 149–159.

Siderits, M. 1997. "Buddhist reductionism". *Philosophy East and West* 47(4): 455–478.

Siderits, M. 2008. "Paleo-compatibilism and Buddhist reductionism". *Sophia*, 47: 29–42.

Siderits, M. 2015. *Personal Identity and Buddhist Philosophy: Empty Persons* (2nd ed.). Aldershot: Ashgate.

Smart, J. J. C. 2006. "Metaphysical illusions". *Australasian Journal of Philosophy* 84(2): 167–175.

Stcherbatsky, T. 1919. "The soul theory of the Buddhists". *Bulletin de l'Academie des Sciences de Russie*, 13 : 823–854; 937–958.

Stone, J. 1988. "Parfit and the Buddha: Why there are no people". *Philosophy and Phenomenological Research*, 48: 519–532.

Stone, J. 2005. "Why there still are no people." *Philosophy and Phenomenological Research* 70: 174–192.

Strawson, P. F. 1992. "Freedom and necessity". In *Analysis and Metaphysics*. Oxford: Oxford University Press.

Strawson, G. 2006. "Realistic monism: why physicalism entails panpsychism", *Journal of Consciousness Studies* 13(10 –11): 3–31.

Strawson, G. 2015. "Self-intimation". *Phenomenology and the Cognitive Sciences* 14: 1–31.

Swinburne, R. 2013. "How to determine which is the true theory of personal identity", in G. Gasser and M. Stefan, *Personal Identity: Complex or Simple*. Cambridge: Cambridge University Press, 105–122.

Thomasson, A. L. 2007. *Ordinary Objects*. Oxford: Oxford University Press.

Thomasson, A. L. 2008. "Phenomenal consciousness and the phenomenal world". *The Monist* 91(2): 191–214.

Venkataramanan, K. 1953. *Sāṃmitīyanikāya Śāstra*. *Visva-Bharati Annals*, 5: 155–243.

Wegner, D. 2002. *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.

Wittgenstein, L. 1973. *Philosophical Investigation*. 3rd ed., (tr.) G. E. M. Anscombe, Oxford: Blackwell.

Zahavi, D. 2005. *Subjectivity and Selfhood: Investigating the Fist-Person Perspective*. Cambridge, MA: MIT Press, 2005.

Zahavi, D. and Kriegel, U. 2015. "For-me-ness: What it is and what it is not". In D. Dahlstrom, A. Elpidorou and W. Hopp (eds.), *Philosophy of Mind and Phenomenology*. London: Routledge, 36–53.