An Irrealist Theory of Self

By Jonardon Ganeri

A false self in the midst ye plant, and make
   A world around which seems;
Blind to the heights beyond, deaf to the sound
   Of sweet airs breathed from far past Indra’s sky
Dumb to the summons of the true life kept
   For him who false puts by.

-Edwin Arnold, The Light of Asia, Book The Eighth.

Neither self nor no-self in reality is to be found;
The Great Sage ruled out the views made of self and of no-self.

– Nāgārjuna, Ratnavālī 2.3

1. Three Versions of the ‘No-Self’ Thesis

It has become a commonplace to read the ‘no-self’ theory of the Buddhist philosophers as a reductionist account of persons. In Reasons and Persons, Derek Parfit himself seemed to endorse the association, having learned of the Buddhist theory from his colleague at All Souls College, Bimal Krishna Matilal.¹ The Buddha’s denial that there are real selves metaphysically distinct from continuous streams of psycho-physical constituents lends itself, to be sure, to a reductionist interpretation. I believe, nevertheless, that there are good grounds for scepticism, and I think it is time for scholars of Buddhism to be more cautious about the identification than they have been up until now. Different Buddhist schools, not to mention different thinkers within particular schools, have given widely varying philosophical construals of the Buddha’s claim.
about ‘no-self’, and, while some thinkers and some schools might favor a reductionist reading of the claim, others, I would argue, do not. In this paper, I will examine the theory of persons of one such, the Mādhyamika Buddhist Candrakṛṣṇa (c. 600–650 C.E.). Candrakṛṣṇa’s interpretation of the “no self” slogan is, I believe, anti-reductionist but irrealist: persons are not reducible to psycho-physical streams, nor are they real existents distinct from the stream. How is it possible for him to say both these things? Let us see.

I begin by charting the terrain. The language of self–use of personal pronouns, proper names, and so forth–is, apparently, representational; that is to say, it appears to refer to and make claims about entities of a certain kind, claims that are assessable as true or false, and whose truth or falsity is determined by the properties of the entities so referred to. Realism about persons is the thesis that appearances here are not deceptive: the terms in this discourse do indeed refer; moreover, assertions made within the discourse are often true, and when true, they are true because the entities so referred to do indeed have the properties ascribed to them. Reductionism has, historically, been a resource of those who would like to defend realism against a perceived threat of ontological proliferation. Reductionism is the thesis that statements in the disputed discourse, when true, are true because of the truth of statements in another discourse, one whose terms refer to entities whose status is less problematic. A committed naturalist who wants neither to admit persons into his or her primitive ontology nor to write off all talk of persons as unintelligible, finds in the strategy of reduction the hope of a salvage operation: the language of self is derivable, with the help of appropriately constrained “bridge-laws”, from the language of psycho-physical continuants. The motto of reductionism is “realism at no extra expense!” Parfit expresses his reductionism in terms of a commitment to the impersonal description thesis, the claim that “though persons exist, we could give a complete description of reality without claiming that persons exist.”

How might reductionism with regard to some domain of discourse be resisted? Broadly, there are three available anti-reductionist strategies. One is to show that the reducing terms (or their referents) inherit their “shape” or principle of identity from the terms being reduced. If there were no means to individuate distinct streams of psycho-physical elements other than with reference to persons, the “reduction” of the language of self into the language of streams would, though formally adequate, fail to be genuinely reductive. A second strategy is to show that the reducing item fails to do the explanatory work of the item being reduced. This strategy seeks to point out distinctive, perhaps non-causal, explanatory work done by the talk of persons or selves. A third strategy hopes to demonstrate that the language being reduced exhibits an essential feature not present in the reducing language. For instance, the anti-reductionist might seek to show that there is an essentially perspectival and consequently subjective
element in our talk of selves not captured within the non-perspectival framework of physical science. These three strategies are respectively exemplified in the anti-reductionist arguments of John Campbell, Richard Sorabji and Thomas Nagel.

Contra reductionist propaganda, an anti-reductionist is not forced into an endorsement of substance dualism. Resistance, however, does incur a substantive obligation: to give an account of the relationship—let us call it the “dependence” relationship—between talk of persons or selves and talk of psycho-physical streams. That obligation is acute for the anti-reductionist who wishes still to be a realist about persons, but it exists as well for the anti-reductionist who prefers to assume an irrealist position. There are two paradigms here for irrealism. One concurs with the realist that the language of self is a language of referring terms, and of claims made true by the properties of entities so referred to. Where he or she departs from the realist is in issuing a denial that there are, in fact, entities of the kind in question. This is an ‘error theory’ of the self: our talk of selves is representational but globally in error. There is nothing in the world for our proper names and personal pronouns to denote, nothing to make the statements about selves true: our discourse about persons is systematically mistaken. Terms like “I” and “you” are empty terms (and perhaps we can remain agnostic here about whether the statements in which they occur are all false or all neither-true-nor-false). The other irrealist paradigm rejects the assumption shared by the realist and error-theoretic irrealist, the assumption that the language of self is genuinely representational. Both have been misled by the surface grammar of this language; in fact, statements about persons are not truth-apt at all, nor is their function to refer and make claims about things referred to. What is needed, say these “non-factualist” irrealists about the self, is an explanation of the way we use the language of self freed from the mistaken assumption that it is a species of representational discourse; just as, in the prescriptivist and quasi-realist traditions in ethics, the role of ethical statements is not to assert anything but to express a moral attitude and prescribe against specified modes of conduct.3

Even from this brief review of the terrain, it is clear that there are three distinct positions that a theorist working under the banner of ‘no-self’ might strike up. To be sure, he may be a reductionist, in which case the doctrine of ‘no-self’ is a thesis that there is no sui generis entity irreducible to a psycho-physical stream. But he could also be an error-theoretic irrealist, in which case he will read the slogan as a strict denial that there is anything to which the representational discourse of self refers; he could be a non-factualist irrealist, opting to deny that the surface grammar of our talk of self is a fair guide to its true function. Each of the three positions is, to some extent, in conflict with the pre-theoretic common-sense view of the self. Of the three, the first, reductionism, is the least revisionary. The reductionist’s revolution leaves everything pretty much as it was before, subject to a little
ontological springcleaning; but it is vulnerable to any of the anti-reductionist strategies outlined above. The error-theorist is more revolutionary, for if the entire language of self rests on a massive mistake, surely in an ideal world it should be set aside altogether. The error theorist owes us an explanation of how such a mistake came to be made, how talk of self can have a utility even if it is so colossally mistaken, to what extent thought and talk about the self can be set aside, and what would be the consequences of doing so. In particular, the error-theorist might favor replacement, the substitution of the language of self with talk only of streams or of entities that are by hypothesis reducible to them. Suppose we define the “language of stream-selves” to be the richest language that is, by hypothesis, reducible to the language of streams. Then Parfit's most recent position, if I have understood it rightly, is a combination of an error-theory about the language of selves with a recommendation that we substitute this language for the language of stream-selves.4

The non-factualist owes us an explanation of another kind. If the function of our talk of self is, in spite of appearances, not to talk of selves, then what does it do, and how is it related to the talk of psycho-physical streams? The non-factualist will need to explain the extent to which the intuitive commitment to representationalism can be given up, but will distinguish that question from a further one about the extent to which the language of self, irrealistically construed, can and should be surrendered.

2. Candrakṣārti’s Sense of Self

Candarṣārti is a non-factualist. He denies that persons are identical to psycho-physical streams, and he also denies that they are distinct; that is, he rejects both reductionism and substance dualism. He argues that there is a point and a function to the “language of self” which is not to refer to and make claims about selves; that is, he rejects representationalism. He discusses the “dependence” of the notion of self on the psycho-physical streams, and he speculates on the possibility for, and consequences of, the giving up of our pre-theoretic commitment to representationalism in the domain of discourse. The principal textual evidence for these claims comes from chapter 6 of the Introduction to the Middle Way (Madhyamakāvatāra),5 and from his commentary to chapter 18 of Nāgārjuna’s famous Lead Verses on the Middle Way (Mālamadhyamakakārikā).6 If we are to establish the exact nature of Candarṣārti’s view, it will be worth our while to review the texts with some care.

Candarṣārti cautions against a reading of the ‘no-self’ doctrine that equates it only with the rejection of the “classical” theory of self as an eternal, substantial, independent entity. That would trivialize the doctrine, whereas in fact its implications are much more profound:

[Let us suppose for the moment that] when the absence of self is understood,
[this simply entails] a rejection of this “eternal self.” But this [refined concept of an eternally existent self] is not considered to be the basis of the clinging to an “I,” and therefore why would the philosophical view of a real, substantial self be supported by understanding the absence of a self [in this manner]? Such a proposition would be marvellous indeed!

[It is as if] someone were to see that a serpent had taken up residence in a hole in the wall of his home. He proceeds to assure himself that there is no elephant in the house, and by doing so, he manages not only to dispose of his fear [for the imaginary elephant], but he also rids himself of any apprehension for the serpent! Indeed, our opponent is strikingly naïve [if he would hold such a position]. [MA 6.140–141; trans. Huntington].

Indeed, the idea that “I” refers to a substantial self is not true even at the level of everyday convention (saµv®ti-satya). It is not even a self-deception: this false self is but false theory –

A self like this simply does not exist, for it is no more produced than is the son of a barren woman. Moreover, it makes no sense that it should serve as the basis for clinging to an “I”: We do not consider it to exist even from the perspective of the [truth of the] screen. [MA 6.122; trans. Huntington].

Against the conception of self as mental substance, Nâgârjuna had already said that “if [the self] were something other than [the psycho-physical stream], it would not be characterizable in their terms” (MK 18.1cd). That is to say, we would not be able to say things like “I am hot,” “I am walking,” “I am happy.” Candrakârti explains:

If the self were other than the psycho-physical elements, its definition would not mention them. The five psycho-physical elements are defined as bodily form, experiencing, seizing on the specific character of things, shaping one’s dispositions, becoming aware of objects. The self imagined as wholly other than the psycho-physical elements, just as consciousness is other than physical form, would be require a separate definition. [PP 18, 343–344].

If the self is a substance wholly other than the constituents of the psycho-physical stream, then it will have to be described in terms exclusively appropriate to it, just as the Cartesian dualist describes mind in one set of terms (“thinking”, and so forth) and matter in another (“spatial extension” and so forth.). What then is the origin of the false conception of the self as a distinct substance?

They who, from fear and by not comprehending the nature of acquisitive reification (upâdâya-praâjñapti), fail to understand that the self is merely nominal, who have veered away even from the truth of concealment, who are deceived by false thinking (mithyåkalpanå) into what is only an apparently good argument (anumånåbhåsamåtra), in their delusion conceive of a self and enunciate a definition.... They who seek freedom consider acquisitive reification to constitute the ground of attachment to self among those who run after error (viparyåsa) and false belief (avidyå). [PP 18, 345].

Therefore there is no self different from the psychophysical aggregates, for apart from the aggregates it cannot be established. Nor is it considered to be the cognitive basis for clinging to an “I,” which is a part of everyday
An Irrealist Theory of Self

This philosophical view of a self is unreasonable. Confusion about source of second quote and its source relation to third. Even those who have wandered for eons as animals do not perceive this eternal, unborn [self], yet we can see that they still cling to an “I.” On this account, there is no self different from the aggregates. [MA 6. 124–5; trans. Huntington].

The conception of self as a distinct substance is a false theory produced by the reification of the facts about the first-person and the sense of self.

If the ‘no-self’ doctrine is a rejection of the conception of self as mental substance, it is also and equally a rejection of a conception of self as reducible to the psycho-physical stream, the aggregate of the five skandhas. Against reductionism, Nāgārjuna has already stated that “if the self were identical with the psycho-physical stream a part would be rising and [a part] falling” (MK 18.1 ab; cf. 27.6). Candrakṛtī’s explanatory comment is that:

The meaning is that one who thinks that the self is the psychophysical stream arrives at a self as that which rises up and that which falls away, – “a part rising and [a part] falling” –, because the psychophysical stream has a part rising and a part falling. However, this is most undesirable, for it implies the fault that the self is plural (ātmanekatvado?a). [PP 18, 342].

The argument that there would be a division of subjecthood if reductionism were true is restated in his further discussion of the topic in MA 6.127–37:

If the self is the psychophysical aggregates, then there would have to be a plurality of selves, since there is a plurality of aggregates. [MA 6.127 ab; trans. Huntington]

The Buddha, it is true, spoke of the self as the aggregates, but that was only in order to counter the false view that the self is something other than the aggregates [MA 6.132], and in fact:

The self is similar to a carriage, and the quality of being a carriage derives from the assembled composite of its parts. However, in the śāstras it is said that the self is merely dependent on the aggregates, and on this account the self is not to be directly equated with the composite of the aggregates. [MA 6.135; trans. Huntington].

Candrakṛtī’s reason for distancing the Buddhist texts from a reductionist account of self is that he does not regard this account as giving a proper analysis of the ordinary concept of self; the latter, he will say, is best analysed in terms of the notion of “appropriation” (upādāna), and:

It is inherently unreasonable that the appropriator and the appropriated substratum are identical, for if this were the case, then the ‘object of action’ and the ‘agent’ would be identical as well. [MA 6.137 ab; trans. Huntington].

vol.XII no.1  2004  THE HARVARD REVIEW OF PHILOSOPHY
If the self is reducible to the psycho-physical elements, then no distinction can be made out between the activity of appropriating, and the things appropriated; but our conventional conception of self is precisely that which appropriates the psycho-physical elements to itself.

If the first-person does not refer either to a distinct mental substance or to the psycho-physical aggregate, what alternative is left? In fact, there is a perfectly adequate way to explain how the first-person is used and in what our everyday conception of ourselves consists in without the hypostatization and reification of self: one’s sense of self consists in the appropriation of psycho-physical elements to oneself. Candrakṣṛti argues against any representationalist construal of the language of self:

Consequently, the basis of clinging to an “I” is not an entity. It is not different from the psychophysical aggregates, it is not the essence of the aggregates, it is not the receptacle of the aggregates, and it does not possess them; [it] is established in dependence on the aggregates. [MA 6.150; trans. Huntington].

In what sense does the concept of self “depend on” the psycho-physical stream, if it is neither reducible to it nor wholly other than it? The leading metaphor for this relationship of dependence is the nature of fire’s dependence on fuel. Nāgārjuna again provides the lead, saying “Everything expounded in terms of fire and fuel is, without exception, applicable to self and the psychophysical aggregates” (MK 10.15). Candrakṣṛti comments:

What the self possesses is what is appropriated (upādāṇa), namely, the five appropriative factors of personal existence. What is commonly thought of as being based on these factors is the appropriator, the conceiver, the active agent and this is said to be the self. Because the “I-me” sense (ahamkāra) is made into an object, the illusion of the “I” is conceived as in and of personal existence. The argumentation concerning the self and what it possesses is to be understood as exactly parallel to that expounded for fire and fuel. [PP 10, 212–213].

This, clearly, is what Candrakṣṛti considers the everyday conception of self to consist in, an appropriative act of laying claim to the elements in one’s own psychophysical aggregate, an act that does not require there to be any ‘entity’ or ‘object’ that is the self, nor any of the usual apparatus of reference to things:

Because it is taken for granted in the context of everyday experience, we consider the self also to be the appropriator, in dependence on the psychophysical aggregates, the elements, and the six sense organs with their respective objects. The appropriated substratum is the object of the action, and the [self] is the agent.

However, because there is no such entity, it is neither eternal nor transitory; it is not produced, nor is it destroyed. It has no quality of permanence and so forth, nor of identity, nor of difference. [MA 6. 162–3; trans. Huntington].
In the Ratnāvalī, Nāgārjuna avails himself of another useful metaphor:

Just as through the medium of (upādāya) a mirror one sees the reflection of one’s own face, even though it is in fact nothing real, even so one reaches a sense of self through the medium of the psycho-physical elements, though in truth it is no more real than the reflection of one’s face. Just as without the medium of a mirror, no reflection of the face can be seen, even so without the medium of the psycho-physical elements, there is no sense of self. [R 1.31–3; translation Tucci].

I will continue to explore Candrakṣaṇṭi’s theory of the self conceived of as an appropriating to itself the constituents of the psycho-physical stream—and therefore as “no-thing”—in the next section. Is it really possible, though, that the misconstrual of “I” as referring to some thing can be eliminated, that we can rid ourselves of this self-inflicted self-deception, of this false self? According to Candrakṣaṇṭi, it is indeed possible. His starting point is again a verse from Nāgārjuna’s Ratnāvalī:

The psycho-physical complex originated from the sense of self, but this sense of self is in reality false (anatā). How can the sprout be true when the seed is false? [R 1.29]

Candrakṣaṇṭi interprets this passage as referring to the conception of “I” as an object, and explains that this sense of “I” is merely an optical illusion; it is the way the “I” appears when viewed, as it were, from a distance:

The sun, at the end of a summer’s day when it is throwing out fiery rays of light and just as it enters that part of the heavens where there is no cloud, emits slanting rays like elongated sparks from a blazing fire and warms the dry earth beneath. If one is in the vicinity of this dry area a visual illusion gives rise to a mirage which seems to be water. For those at a distance it seems to be clear blue water; but for those close by it does not give rise to a mirage.

Similarly, for those who are far removed from viewing the nature of self and own as they really are, who are caught in the cycle of birth and death, in the grip of the misbelief of primal ignorance, for such, a false thing—the self as hypothesized on the basis of the skandhas—manifests itself as real. But for those close by who see the truth of these matters, no such false thing manifests itself. [PP 18, 347; translation Sprung].

But linked to this, continuously and strongly, beings cling to “I”, and all that “I” possesses is conceived as “mine”. This self will manifest empirically, the fruit of ignorance, as long as it’s not subject to analysis. Without a worker, there’s no work performed. And likewise without “I” there is no “mine”. Perceiving that both “I” and “mine” are void, the yogī will be utterly free. Vases, canvas, bucklers, armies, forests, garlands, trees, houses, chariots, hostelries, and all such things that common people designate, dependent on their parts, accept as such. For the Buddha did not quarrel with the world. Parts and part-possessors, qualities and qualified, desire and those desiring, defined and definition, fire and fuel—subjected, like a chariot, to sevenfold analysis are shown to be devoid of real existence. Yet, by worldly, everyday convention, they exist indeed. [MA VI 164–7; trans. Huntington].
The misconception that “I” is a referential expression is an illusion which results from standing too far back from the facts of personal identity, namely the psycho-physical stream, just as a cloud or a table looks from a distance like a solid object, but dissolves into vapour or atoms on closer inspection. To put matters another way, the level of description to which this use of “I” belongs is a level that concerns itself with such matters as the re-identification of persons over time, rather than the level of description at which one is concerned with matters of composition. These two levels of description are not incompatible with one another, but when it comes to settling questions about what there really is in the world, the ‘view from close up’ is the proper one to assume. We can certainly agree with Candrakṣṭiti that the world described according to the ‘view from close-up’ need not share the ontology of the world as described from far-off. What requires further argumentation is the additional claim that the ontology of the ‘view from close-up’ has a greater claim to reality than the ontology of the view from far-off. We do not typically regard clouds and tables as optical illusions simply because they disappear when we ‘zoom in’. The same is true of the chariot—why does the possibility of analysis, or the redescription of the chariot in terms of its parts, call into question the reality of the chariot itself, described at the level of wholes? Why should there not be, as philosophers as diverse as Michael Dummett and the Jainas have maintained, different but compatible levels of description, each with its own proper domain of objects? Why should we not say that when we look at an impressionist painting close-up and see only dots of colour, the disappearance of the painting is the illusion and the “real” painting is the one seen from far off?

According to the non-factualist, we have simply misunderstood the grammar of “I” if we think of it as a referring term, rather than a term whose linguistic use is to perform an expressive act of appropriation. Suppose then we rid ourselves of at least this mistake. Can a human person go further, and give up the language and concept of self altogether, even understood as having the non-representational use the non-factualist claims it to have? Nāgārjuna:

Without ‘me’ or ‘mine’ because the self and what belongs to it are still, he who is without ‘me’ or ‘mine’ does not exist. One who perceives that which is free of ‘me’ and ‘mine’ does not perceive. When ‘me’ and ‘mine’ are destroyed both within and without, appropriation (upādāna) comes to an end; with its demise, rebirth ends. [MK 18.2  CD – 18.4]

Nāgārjuna’s startling claim would appear to be that with the cessation of all use of the vocabulary of self, the person quite literally ceases to exist. Candrakṣṭiti modifies the claim, interpreting it as a shedding of the misconception that there is a permanent substantial self, something that is possible for the adept at a final stage of the Buddhist path. But if all activity of self-appropriation of the psycho-physical is abandoned,
then, he too agrees, it is the end:

This belief in the permanent self is brought to an end by longer having a sense of self and of what belongs to self. From that sense coming to an end the four kinds of appropriation (upādāṇā) – to sense pleasure, to dogmas, to morality, and to belief in the permanent self – ceases. From the cessation of appropriation, personal continuity defined as re-birth is ended. The sequence of stages in the cessation of rebirth is like this. Appropriation having ceased, there is no personal continuity. When personal continuity has come to an end, how can there be the cycle of birth, old age and death? [PP 18, 349 on MK 18.5a].

There are two stages in the path from ignorance about the self to transformation of mind. One stage involves the elimination of a false representational conception of self and the language of self, a conception that is to be replaced by an appropriative model of self-knowledge and self-reference. The second stage involves the elimination even of acts of self-appropriation. While the completion of the first stage frees one from egotism and self-interested or self-centered motivation, but leaves one with a residual sense of self and an ability to discriminate between oneself and others, the completion of the second stage is the culmination of a process of complete self-annihilation, self-surrender and loss of autonomy. This is, perhaps, nirvāṇa, and it is for that reason that the way of the bodhisattva is the way of the first stage alone. The bodhisattva retains a residual sense of self-sufficiency for moral agency and altruistic action. What I shall therefore be interested in is the possibility of a non-vacuous conception of self that is free of the representationalist error – a sense of self without a false self.

3. Of Me and Mine: The Appropriativist Theory of Self

If I am right, Candrakīrti’s view is that the language of self – use of words like “I”, “mine”, “you”, and so forth – is not properly understood as having a representational function. This is an illusion of common sense and surface grammar, an illusion which dissolves on close analytical inspection. Giving up this illusion implies much more than merely giving up the idea of a mental substance; it requires surrender even of a reductionist conception of self. But it does not yet follow that there is no use-explanatory account of the language of self. ( “Use-explanatory” feel awkward–perhaps “no explanatory account of the utility of the language of self” or some similar revision?) What, then, is the use of the language of self? The use of “I” serves, it seems, an appropriative function, to claim possession of, to take something as one’s own (compare upādāṇa: “the act of taking for one’s self, appropriating to one’s self”). (This preceding sentence is a little unclear, because the “use of something serving a particular function” is hard to follow; perhaps just ‘ “I” serves’ or ‘The word “I” serves’) The appropriation in question is to be thought of as an activity of laying claim to, not the making of an assertion of ownership. When I say, “I am happy,” I do not assert ownership of a particular happy experience; rather, I appropriate the experience within a stream,
and in doing so lay claim to it. Call this the appropriativist theory of self. It has the virtue of elucidating the relation of dependence between the language of self and the psycho-physical streams, and it ties in with the sorts of consideration about the grounding of our concept of self that have been brought to the fore by Richard Sorabji. In particular, Sorabji argues that facts of ownership are the “further facts” not accounted for by a reductionist theory of self. The point of contention between the realist and the irrealist is over whether to give a deflationary or a robust interpretation to the “further fact” of ownership.

Öryadeva, the younger contemporary of Nāgārjuna and co-founder of Madhyamaka Buddhism, rehearses a curious argument in support of the Buddhist thesis that there are no selves:

That which is self to you is not self to me; from this fixed rule it follows that it is not self. Indeed, the construction [of self] arises out of the impermanent things. [Cf 10.3]

T.R.V. Murti ventures the following paraphrase of this difficult verse:

If the ātman were a real entity, there should be agreement about it. On the contrary one’s (self) ātman is anātman (non-self) for another, and vice versa; and this should not be the case if it were an objective reality.

Murti’s paraphrase hints at an interesting reading of the argument in the verse. Why do human beings have a concept of self? What work does it do? Possession of a concept of self is important because it enables me to think of my ideas and emotions, my plans and aspirations, my hopes and fears, as mine—as belonging to me and the proper objects of my concern. Equally, it is the concept of self which is in play when I think of your ideas and emotions, your plans and aspirations, your hopes and fears, as yours—the proper objects of your concern. A human being without a concept of self, therefore, would be a seriously impoverished creature. Lacking the concept of self, I would not be able to draw the distinction between what is mine and what is yours, and, unable to make that distinction, I should also lack the capacity to form plans or act on intentions, not to mention the ability to make promises, enter into commitments, or accept responsibility. For example, I could not be in a position to intend to do something, for I would not understand that the intention is fulfilled only if I and nobody else performs the action intended. It is our capacity to make out that distinction to which Öryadeva draws our attention when he says that “that which is self to you is not self to me.” What we have here is an adequacy condition on potential theories of self.

It is at this point that Öryadeva’s argument takes an interesting turn. What Öryadeva next argues is that the classical theory of self fails the adequacy condition. The classical theory asserts that selves exist as real and permanent entities; it reifies the facts of selfhood, and accounts
for our possession of a concept of self on the model of our possession of concepts of external objects. Ørydeva’s refutation of the classical theory involves the claim that this is a false model. For our concept of an object is a concept of something public, an inhabitant of a shared world, something that can equally well and simultaneously be the common focus of your and my attention. To reify the self, to explain our possession of a concept of self on the model of our possession of the concept of an object, is thus precisely to render one incapable of explaining why we have the concept of self in the first place, that it sustains the notion of something being exclusively or distinctively mine.

Candrakṣṇi’s commentary on the first half of CÍ 10.3 strengthens the argument and relates it to the Buddhist theory:

It follows that the self does not exist essentially (svarāpata?). If the self were to exist essentially, then just as it would be the foundation of one person’s sense of “I”, so it would be the foundation of everyone’s sense of “I”. For it is not the case that the essential nature of fire is burning and yet that sometimes it does not burn. So if the self were to exist essentially, it would be the self for everyone and the focal point of their sense of “I”. And this is not the case, for That which is self to you is not self to me; from this fixed rule it follows that it is not self.

That which is self to you, the focal point of your sense of “I” (ahākāra) and self-concern (atmasneha), that indeed is not self to me; for it is not the focal point of my sense of “I” and self-concern. This then is the fixed rule from which it follows that it is not [a real thing]. There is no essence to such a self as is not present invariably. One should give up the superimposition of a self, as being something whose content of which is unreal (asadartha).

This seems to be an inversion of the third type of anti-reductionist argument mentioned above: there are features of the language of self that the realist construal of that language necessarily omits.

4. Body-swapping and Personal Survival

Further discussion of the argument, or at least a closely similar one, is available in the Mahāprajñāpāramitāśāstra. I will not follow Étienne Lamotte in endorsing the Chinese tradition’s attribution of the text to Nāgārjuna: but whoever is responsible for the original compilation, it remains an important source for early Madhyamaka. The discussion begins with three suggestions as to why a notion of self as distinct from psycho-physical streams is required. The first is that “each individual person conceives the notion of the self in relation to his own person (svakāya), not in relation to that of someone else. If, therefore, he wrongly considers as self the non-self of his own person, he would also wrongly consider as self the non-self of another person.” (Citation) Lamotte reads the argument as an insistence that my concept of myself as distinct from you is explicable only if the concept “corresponds with something real,” and regards it as presenting the position refuted in CÍ 10.3. The second argument is that “if there is no internal self, [being given that]
acquaintance with colours arises and perishes from moment to moment, how does one distinguish and recognise the colour blue, yellow, red or white?” (Citation) This looks as if it is an early statement of what was to become a very influential argument in favour of the self, found in the commentaries to Nyāya-sūtra 3.1.1, that the self is required by the possibility of psychological unity and the reidentification of objects. Finally, another argument familiar from the Nyāya-sūtra is rehearsed, that “if there is no self, and since the knowledge of human activities, arising and perishing repeatedly, all disappear with the life of the body, who then is left to reap the rewards—good or bad? Who endures sadness or happiness? Who is set free?” That echoes the familiar argument of Vātsyāyana in his commentary to Nyāyasūtra 1.1.10.

Our present interest is in the first of these arguments. If the concept of self is grounded in notions of self-concern and a sense of “I” as distinct from “you”, then the question we need to consider is whether what matters to me when I think about my survival and future well-being is the continuity and status of the stream of psycho-physical events, or whether what matters to me is the future existence and status of a person numerically identical to myself. Derek Parfit famously introduced a science-fiction example to test our intuitions about this question. He asked us to imagine that we are stepping into a teletransporter, a machine that completely destroys the body at the same time as it transmits all the physical and psychological data necessary to reconstitute the person in another transporter remotely situated, and he argues that we care less about the imminent destruction of our old body than about the continuity of our psychological lives in the life of the replica. What about the slightly modified case in which, as a result of a malfunction, my old body is not destroyed but continues to live as before (Parfit calls this the Branch Line case)? Suppose I step out unharmed from the teletransporter, only to be told that I am about to die. Do the well-being and good prospects of my replica far away (someone who has all my memories and personality traits, and calls himself by my name) provide me with comfort in the face of my imminent death? Here our intuitions are less clear, and for Parfit that is exactly how they should be if our concern tracks the continuity of the stream rather than numerical identity. The relation of numerical identity is definite, relations of psycho-physical continuity indefinite; and if there is an indefiniteness in the object of our self-concern and sense of “I”, then these notions follow continuity relations rather than relations of numerical identity.

The author of the Mahāprajñāpāramitāśāstra brilliantly anticipates Parfit here. He claims that there are circumstances in which ambiguity creeps into the notion of self, and he tests our intuitions, not by means of science fiction but with stories taken from the legends. He tells the story of ...

... a man who has been charged with travelling far finds himself spending the night alone in a deserted house. In the middle of the night, a demon carrying
on his shoulders a dead man came and put it down in front of him. Then another demon came in pursuit of the first demon, and angrily reproached him, saying: “That dead man belonged to me—how come it was you who brought him here?” The first demon replied: “He is my property; it is I who brought him myself.” The second demon responded: “It is in fact I who brought the dead man here.” These two demons, each grasping the dead man by a hand, argue over him. The first demon says: “Here is someone we could interrogate.” And the second demon starts to question him and says: “Who brought the dead man here?” The man comes up with the following reflection: “These two demons are very strong. If I tell the truth or if I lie, my death is certain, and in either case I will not be able to avoid it. What good is it to lie?” He declared therefore that it was the first demon who had brought the dead man there. At that point, the second demon very angrily seized him by the hand and tore his hand off and threw it to the ground. But the [first] demon took a [hand] from the corpse and he fitted it to him. In the same manner, he substituted in the man’s body the two arms, the two feet, the head and the sides of the corpse. And together the two demons devoured the body of the man whose body they had substituted with that of the corpse. Then after wiping their mouths, they left. The man reflected again: “I have seen with my own eyes those two demons devour completely the body which was born from my mother and father. Now my present body is entirely made up of the flesh of another. Have I now a body or have I no longer a body? If I believe I have one, it is entirely the body of another; if I believe I don’t have one, here nevertheless is a visible body.” When he thought about this, his mind was greatly troubled, and he acted like a man who had lost his reason. The next morning, he left and set out again. Arriving in a kingdom, he saw in a Buddhist stupa an assembly of monks, from whom he inquired if his body existed or not. The monks asked him: “Which person are you?” He replied: “I do not know if I am a person or if I am not a person.” He then told the assembly what had happened. The monks said: “This man has by himself recognised the non-existence of the self; he will be saved easily.” Addressing him, they said to him: “Your body, from its origin right up to today, has always been deprived of self. And that’s not just the case now. It’s simply because the four great elements were combined together that you thought ‘This is my body’. Between your past body and the one of today, there is no difference.” The monks converted him to the Path; he severed his ties and became an Arhat.

The moral of this story is three-fold. First, the recognition that I have a concern for my future survival and a concern for my own well-being, as distinct from any concern I may have for your survival and well-being, does not require the postulation of selves as objects (‘false’ selves). Indeed, if Óryadeva is right, such a postulation would singularly miss its mark, trying, as we might put it, to render objective the essence of the subjective. Second, the preservation of physical continuity is not the only matter of importance in the question of my survival; psychological continuity without physical continuity is sometimes enough to assuage a concern about the future. Third and finally, it is far from certain and definite whether I attach more weight to my physical continuity or to my psychological continuity. Both seem important, and cases where they come apart are extremely vexing (especially if one is the victim of a body-snatcher). This very indefiniteness supports the view that it is relations of continuity rather than of strict numerical identity which
are in brought into play in the application of notions of self-concern and sense of “I”. The author of the Mahåprajñåpåramitåßåstra did not, but easily could have continued his story by imagining that the demons regurgitated and reconstituted the body of the unfortunate man. In this Ship-of-Theseus scenario, just as in the Branch Line version of the teletransporter, our intuitions become blurred.

5. Óryadeva Again

Let us return now to the second half of Óryadeva’s argument in Catuḥätaka 10.3. Óryadeva has presented us, in effect, with an adequacy condition on any putative theory of self—that it must be able to explain the possibility and significance of the contrast between mine and yours—and he has argued further that the classical theory of self, which claims that selves are real mental entities, is unable to explain even how that contrast is possible. What then about the Buddhist theory of self itself? How well does it fare? The Buddhist theory, as articulated by Candrakṛṣṭi in his comment on the second half of the verse, is that a concept of self is a superimposition onto the aggregate stream of psycho-physical events:

If there is no self, whence this sense of “I” and self-concern? Our author says:

Indeed, the construction [of self] arises out of the impermanent things. Although, from the rule mentioned above it follows an actual self as essential and distinct from the psycho-physical constituents never exists, still a constructed idea of self [arises] out of the impermanent things that do exist—physical attributes, cognitions, sensations, volitions and conceptions (rëpa-vedanå-saµjñå-saµskåra-vijñåna). A superimposition (kalpanå) is made whose content is unreal (abhåtårtha), that the self exists and lives and moves about. A self is hypostasized (prajñapyate) as dependent on the psycho-physical constituents, just as fire is dependent on kindling. The meaning of the statement that there is a construction of self out of the impermanent elements is that it is constructed by hypostasization dependent on the notion that it does not exist essentially but as different in nature from the constituents and yet determined in a fivefold way.

If it is a mistake to think of a person’s identity as consisting in their possession of an essence, it does not follow that we do not have a concept of self and a notion of self-concern. Candrakṛṣṭi claims that these notions arise out of and bear upon relations of psycho-physical continuity rather than numerical identity over time. I have concern for the future states of the stream of psycho-physical events that is me, you for yours. The contrast between our respective concerns is made out without resort to the thought that what I am is an entity with a strict identity across times. What Candrakṛṣṭi claims, therefore, is that when I express a concern about my future survival or the success or failure of my plans and intentions, what properly sustains my concern is the psycho-physical stream, and not the future condition of a numerically identical self.

A different worry now is that while the Buddhist theory can
make sense of the possibility of the distinction between mine and yours, it is incapable of sustaining a plausible explanation of the point and significance of that distinction. Richard Sorabji has recently pressed the objection against both Parfitian and Buddhist accounts of the self. The substitution of our ordinary talk about myself and other persons with talk about this or that stream of psycho-physical events deprives our use of such notions as responsibility, commitment, credit, blame, pity and compassion, and even intention, shared attention and social referencing of much of their point. Suppose, for example, that I intend to shout at someone in order to draw their attention to me. Substituting the vocabulary of “I” and “they” for talk of this and that stream, we shall have to redescribe the situation as one in one stream contains a desire that another stream “should” contain an attention produced by the shouting. Here, Sorabji says, the word “should” means that it would be desirable, but we are not supposed to talk of anybody for whom it would be desirable, only of the desirability of one stream containing a shouting and the other a resultant attention, presumably because the total situation with its various streams would be more desirable from a rather abstract point of view. But intention, conceived this way, seems to have lost much of its point and motivation precisely because there is no one for whom the outcome would be desirable.” Likewise with credit and blame—we can no longer say that there is a person who deserves credit or blame, for “it would be the act that deserved credit and blame, and the resulting stream, but in the different sense that it would be more admirable, or less so, just as a sunset may be admirable, without anybody deserving credit or blame.” The point is that it no longer seems to matter much whether a particular good experience is included in this stream rather than that: “It is better,” Sorabji continues, “that the universe should contain good experiences rather than bad, but as to which stream of consciousness they might enter, why should that matter? Perhaps because it is preferable that experiences should occur in some sequences rather than others, since their significance will be altered by the sequence. But this would only motivate a preference for certain types of sequence over others. Detachment would have been achieved, but at rather a high price.”

Sorabji’s argument is closely related to what Parfit calls the Extreme Claim, the claim that “if the Reductionist View is true, we have no reason to be concerned about our own futures.” Ordinarily we think that while knowing that somebody will suffer a pain gives me at least some reason to prevent it, knowing that it is I who will suffer the pain provides an ‘additional reason’ to act. According to the Extreme Claim, however, a Reductionist must deny that there are any such “additional reasons”; thus Parfit: “That some pain will be mine does not, in itself, give me any more reason to prevent the pain.” If this is thought implausible, then perhaps, instead, the fact that the pain will belong to the same psycho-physical stream as the expectation itself provides a reason for special concern: this is what Parfit calls the Moderate Claim.
Parfit acknowledges that the idea that there is a distinction between the concern one has for the pain of another and the anticipation of one’s own future pain, and that it is hard to see how a Reductionist can give weight and importance to that distinction, but declares himself unable to find a conclusive argument for or against either the Extreme or the Moderate Claim. Perhaps we should just give up anticipating our future pains.

Óryadeva, clearly, is sensitive to this worry, and the second half of his argument is an attempt to address it. The objection derives its force from the presumption that talk of this and that stream cannot “sustain” or “support” the common use of a vocabulary of me and you, and the whole system of concepts co-implicated with the concept of self, but has rather to be thought of as a substitute for it. But Óryadeva affirms that the popular notion of “I” is constructed out of the impermanent psycho-physical events, and Candrakṛtī explains the relation of dependence between the two vocabularies as akin to the dependence of fire on kindling. Candrakṛtī’s position is that as long as we are careful to separate the notion of self from any false imputation of ontological commitment to a real mental substance, the residue is a concept of self that can be sustained by talk of this and that psycho-physical stream, and in turn sustains the vital contrast between “mine” and “yours.” He is not a reductionist, for he does not claim that the language of self can be translated without loss into the language of psycho-physical streams, any more than talk of the fire is simply talk of the kindling. Nor is his a replacement account, for he does not think either that the use of the vocabulary of mine and yours can be dispensed with entirely, in favour of speaking only about the streams. We need the language of self to make sense of the importance of the distinction between mine and yours, and these are “further facts” over and above but not independent of facts about the streams.

The utility of “I” consists in its being the means by which appropriation occurs, the taking of something as a distinctive object of concern. An appropriative explanation of the use of “I” does not require that it have a representative function, and the false if common-sensical idea that there is a real something to which “I” refers can with effort be given up. Anscombe seems to have a sort of appropriative explanation in mind when she says that the proposition “I am this thing here” is not an identity but a “subjectless” construction meaning “this think here is the thing of whose action this idea of action is an idea, of whose movements these ideas of movement are ideas, of whose posture this idea of posture is the idea.” And Locke seems to be on the same track when he says that we create a self insofar as we reconcile and appropriate actions (Essay 2.27.26), and insofar as our consciousness joins itself to and is concerned for certain bodily parts (2.27.17-18).

What is still far from clear is whether one could go a step further, and give up the appropriative practice while continuing to be a human being in the world. Entirely giving up that system of appropriative concepts would utterly transform one’s inner world, in ways we might
find almost unintelligible. It is hard even to imagine how the mind of a Buddha might work, and perhaps we should not try. It is interesting, though, to make the comparison with the absolute self-surrender expected of a “revolutionary”:

The revolutionary is a lost man; he has no interests of his own, no cause of his own, no feelings, no habits, no belongings; he does not even have a name. Everything in him is absorbed by a single, exclusive interest, a single thought, a single passion – the revolution. In the very depths of his being, not just in words but in deed, he has broken every tie with the civil order, with the educated world and all laws, conventions and generally accepted conditions, and with the ethics of the world. He will be an implacable enemy of this world, and if he continues to live in it, that will only be so as to destroy it the more effectively. For him, everything that allows the triumph of the revolution is moral, and everything that stands in its way is immoral.

The enlightened mind is a revolutionary mind, and the Buddha indeed would have us all become revolutionaries in thought, enemies of the conventional world of pain and suffering. And—who knows—perhaps he did also desire quite literally to create a band of revolutionaries that might upturn an unjust and inflexible brahminical social order, manufacturing the revolutionary mind and using religious soteriology as the well-chosen skilful means, the right clothes in which to wrap a concealed truth. Be that as it may, what I have attempted to show is that there is in Buddhism a viable irrealist and non-reductionist analysis of our ordinary conception of self. This unpacking of the deepest roots of our thinking about ourselves and our place in the world is, I hope it will be generally agreed, a philosophical project par excellence.

NOTES

2 Parfit 1984: 212.
3 A third paradigm for irrealism is, of course, the semantic anti-realism of Michael Dummett; but it is less than clear to me that this paradigm is available in an account of the language of self. Mark Siderits, however, has recently provided an argument for the viability of an anti-realist construal of the language of self, and he is also willing to attribute the construal to the followers of Madhyamaka Buddhism (Siderits 2003). The merits of his “Buddhist anti-realism” require further investigation, as does Roy Perrett’s interpretation of Madhyamaka Buddhism as a type of “minimalism” (Perrett 2002).
4 Parfit 1999.
5 The text is now extant only in Tibetan. I will follow the translation of Huntington and Wangchen 1989.
6 Sprung’s translation of Candrakīrti’s commentary is dependable; there are several readable translations of Nāgārjuna, of which Garfield 1995, although based on the Tibetan, is arguably the best.
7 Sorabji, unpublished.
8 One difficulty relates to the compound ātmāniyamāt. I have followed Candrakīrti in reading ātmā niyamāt ‘from a fixed rule’; the Tibetan, however, reads ātmā aniyamāt (ma
Jonardon Ganeri

?es phyir) ‘from the absence of a fixed rule’.


10 Le Traité de la Grande Vertu de Sagesse de Nāgārjuna (Mahāpraṇītpāramitāśāstra), translated from the Chinese by Étienne Lamotte (Louvain, 1944), volume 2, pp. 736–750. English translation by Anita Ganeri.

11 In his introduction to volume three of the translation, Lamotte withdraws the identification, giving as his grounds a dating of the author of the Traité as approximately the beginning of the fourth century A.D. John Brough speculates that the erroneous attribution to Nāgārjuna may have begun with Kumārajīva: “There is no doubt whatsoever that the Chinese translator Kumārajīva believed that the Traité was the genuine work of the original Nāgārjuna; and for this belief he had reasonable excuse, since the work is in the main Madhyamaka tradition, and refers to and sometimes paraphrases the Madhyamakārikās. This being so, it is natural to assume that the attribution of the Traité to Nāgārjuna is simply a mistake due to Kumārajīva or his near-contemporaries” (Review of Lamotte, BSOAS, p. 165).

12 Derek Parfit 1984, chapter 10.


15 Parfit 1984: 308.


BIBLIOGRAPHY

Cl = Óryadeva’s Catuvṛatakā. See Bhattacharya.
ClV = Candrakṛṣṭi’s Catuvṛatakāvṛtti, a commentary on Cl. See Bhattacharya.
MA = Candrakṛṣṭi’s Madhyamakāvatāra. See Huntington.
MK = Nāgārjuna’s Mālamadhyamakārikā. See Garfield.
PP = Candrakṛṣṭi’s Prasannapadā, a commentary on MK. See Sprung.
R = Nāgārjuna’s Ratnāvalī. See Tucci.

Nagel, Thomas. “What is it like to be a bat?”. [ref]