

CHAPTER 12

BUDDHIST NON-SELF THE NO-OWNER'S MANUAL

MARK SIDERITS

THE Buddha claimed to have discovered a path to liberation from suffering. In this he was not alone. A number of such paths were developed in early classical India. All share the belief that sentient beings suffer because of ignorance about their true identity: we are trapped on the wheel of *saṃsāra* because we identify with things that are not the true self. What set the Buddha's teachings apart from those of other Indian doctrines of liberation is that the Buddha denied there is a self. On his analysis, our ignorance about our identity consists in our holding that there is such a thing as 'I' and 'mine'. Instead, everything is non-self.

In modern discussions of the Buddha's teachings one sometimes finds it claimed that the Buddha only denied the existence of an empirical self, and not the existence of a self that transcends all possible experience.¹ And it is true that his arguments for non-self typically canvass just the five groups of psycho-physical elements (*skandhas*), all of which are possible objects of sensory experience. When, for instance, the Buddha points out that all the psychophysical elements lack the sort of permanence that a self would require, this proves that there is no self only

¹ Albahari 2007 is a recent instance.

given the additional premise that there is no more to the person than the empirically given elements. And the Buddha seems not to have argued for such a premise. So one might suppose that the Buddha only meant to deny that the self is an empirical element, and not deny outright that there is such a thing as the self. This would bring the Buddha's teachings more in line with the views of some of the Brahmanical schools, such as Sāṃkhya and Vedānta. But this is not how the Buddhist philosophical tradition understood the Buddha's teachings. They took the doctrine of non-self to mean that there is no self *tout court*. What I shall discuss here is their philosophical articulations of this understanding of non-self.

Among the various schools that make up Indian Buddhist philosophy, the most widely accepted understanding of non-self is what I call Buddhist Reductionism. Its formulation involves the introduction of a distinction between the terms 'self' and 'person'.² If we think of possible views as to what the word 'I' might denote, we can say that broadly speaking there are two sorts: those that take 'I' to refer to some one part of the psychophysical complex that is its essence; and those that take 'I' to denote a whole made up of the psychophysical elements. By 'self' we shall here mean the one part that is the essence. And by 'person' we shall mean the psychophysical complex as a whole. Then Buddhist Reductionism can be formulated as the two claims:

1. There is no self.
2. The person is a conceptual fiction, something that is only conventionally and not ultimately real.

Of these two claims, it is the second that requires the most explanation. But I shall begin with a brief discussion of the first.

The notion that the self is some one simple part from among the parts making up the psychophysical complex is the consensus view of those (non-Buddhist) Indian philosophers who are realists about the self. The view that this is what 'I' denotes is not as counterintuitive as it might seem.³ We know that our use of 'I' extends beyond the loss of various psychophysical elements: I believe that I would survive the loss of a kidney or the replacement of my heart; I believe that I once existed lacking certain commitments that I now consider central to my existence; and so on. If, like most classical Indian philosophers, we believe that there is

² This distinction is not always honored; in some cases the term *ātman* (which is usually translated as 'self') is used for both. But making such a distinction is required in order to formulate the thesis of the Pudgalavāda school (about which more below), according to which the self is not ultimately real but the person (*pudgala*) is ultimately real. Where the term *ātman* is used for both, a distinction is still drawn between the sense of 'I' had by ordinary people—which corresponds to the concept of a person—and the philosophers' notion of a simple, eternally enduring *ātman*.

³ Of course, that a given view about persons is counterintuitive would not count as an objection to that view for most of the Indian philosophers in question. Since they held that most people lack knowledge of their true identity, our intuitions about our nature are not to be treated as reliable evidence. Being in the wisdom business, Indian philosophers are quite cheerfully revisionist.

rebirth, this phenomenon becomes all the more striking, since between death and rebirth all my body and typically all my psychological makeup are replaced. It then becomes quite tempting to suppose that there is some one part the continued existence of which is necessary for the continued existence of this 'me'. And since this entity must endure over extremely long durations, it is equally tempting to suppose that it is simple or impartite. For if transitory entities come into and go out of existence through the rearrangement of parts, it follows that simple entities must be eternal. There was considerable debate about how else to characterize the self. Some saw it as both cognizer and agent, others held it to be just the subject of experience; some took it to have moral properties, others thought it is nothing but a witnessing consciousness; etc. But all agreed that it is in some sense simple, and that it endures.

The most common Buddhist argument for the non-existence of the self is the argument from impermanence. The Buddha formulated this argument in terms of a survey of the five kinds of psychophysical element: each being impermanent, none could be the self. Later Buddhist philosophers sought to prove that no existing thing could be permanent. One strategy involved the point that a simple entity could not undergo any sort of change, as this would require that one part of the entity change while another part remain the same, and simples have no parts. But since an enduring thing can bring about a change in other things only by itself undergoing some sort of change, it follows that a simple enduring thing could not be causally efficacious. But, it was claimed, for something to exist is just for it to be causally efficacious. Hence there can be no simple enduring entities, and therefore there can be no self.

If there is no self, though, and nothing about us persists even a single lifetime, why does it seem to us as though we are enduring entities? Buddhist Reductionists answer this question by claiming that the person—the whole constituted by the psychophysical elements—is a mere fiction, albeit a useful one. This is what they mean when they say that the person is only conventionally and not ultimately real: our talk of persons is a mere *façon de parler*, but one that we have come to take altogether too seriously. The defense of this claim is relatively complex and will require careful consideration. What one should keep in mind in the course of this investigation is that it is meant to steer a middle course between, on the one hand, the view that there is some real entity that 'I' denotes, and on the other the utter rejection of our concept of the person. The Buddhist Reductionist view of persons is in some respects compatible with views of the self that have recently been discussed in philosophy and cognitive science (and are represented in other contributions to this volume). For instance, a Buddhist Reductionist might welcome a narrative view of the self (the view that we see ourselves as the authors of self-constituting narratives) as one that gives important insights into how the concept of the person serves a useful function. But they would insist that this convergence should not blind us to the fact (as they see it) that the concept of the person is

useful only up to a point, beyond which its employment leads to existential suffering. The Buddhist formulation of Reductionism is meant to show how we can continue to reap the very real benefits that come with our deployment of this concept while not falling into the difficulties that come with complete acquiescence.

Their formulation and defense of the claim that the person is a mere conceptual fiction is grounded in a thoroughgoing mereological reductionism, according to which all partite entities in our common-sense ontology are only conventionally and not ultimately real. As this terminology suggests, Buddhist Reductionism employs a two-tier ontology, according to which only some of the things we take to be real belong in our final ontology, while others are relegated to the status of mere back-benchers. Among the latter are, in addition to persons, such things as chariots, pots, trees, mountains, fires, rivers, hands, and minds. Our folk ontology contains both entities such as these, and the ultimate parts of which they are composed: in the case of macrophysical objects, atoms;⁴ in the case of complex mental entities, simple psychological events such as pains and occurrent desires. But there is good reason not to allow both the wholes and their parts into our final ontology. If, for instance, we count both the pot and the atoms of which it is composed as equally real, then we must say whether the pot is identical with those atoms or distinct from them. Among the reasons against saying they are identical are the fact that the pot is one while the atoms are many; that it is only when those atoms are arranged pot-wise that they are taken to compose one thing and not just be a many (why the prejudice against the same atoms when they are smeared into several odd-shaped mono-molecular films?); and that it is indeterminate at what point the pot ceases to exist as we remove randomly selected atoms from it one at a time.⁵ Among the reasons for denying that pot and atoms are distinct are the fact that they occupy the same space at the same time, and the fact that the pot has no autonomous causal powers (the alleged effects of the pot turn out to be just effects of the atoms).⁶

There being good reasons to deny that the pot and the atoms are equally real, one or the other must go. And we have reason to suppose that the atoms are real: they possess autonomous causal powers. So the pot must go. Still it would be odd

⁴ The atoms that belong in our final ontology are, of course, the philosophers' atoms, not the so-called atoms of current physics, which are misnamed. Both 'atom' and its Sanskrit equivalent (*paramāṇu*) were intended to refer to impartite physical entities. It is because physicists in the 19th century jumped the gun that what we today call atoms are misnamed. What the real atoms might be we do not yet know.

⁵ The last sort of consideration, appealing to the *sorites* difficulties that arise when whole and parts have the same ontological status, was not among those Buddhist Reductionists appealed to in defending their view about wholes.

⁶ Van Inwagen 1995 and Merricks 2001 both give sophisticated and subtle arguments for mereological reductionism about macrophysical objects. But both resist applying the view to persons.

to suppose that the pot is utterly unreal, like the golden mountain or the son of a barren woman. One who believes there to be a golden mountain will inevitably be disappointed in their schemes, while one who believes there to be a pot will have success in storing their honey. It thus seems sensible to grant the pot a kind of provisional or conventional ontological status as a conceptual fiction, something we take to be real only because of our employment of a useful concept as a part of our going theory. As the term 'reductionism' is commonly understood, a reductionist about pots would be one who holds that a pot *just consists in* atoms arranged pot-wise. The two-tier ontology of Buddhist Reductionism helps us understand what this means: that while pots do not belong in our final ontology, a theory that commits us to their existence turns out to be useful given what is in our final ontology (such as atoms arranged pot-wise) and what our interests and our cognitive limitations are (we need containers for our honey, and we're bad at keeping track of individual atoms). To call pots conventionally real is to steer a middle path between realism about pots and outright eliminativism about pots.

All this can be expressed quite neatly if we introduce two distinct truth-predicates, 'ultimately true' and 'conventionally true', with the former applying to sentences that depict how the ultimately real things are, and the latter to sentences that depict how the conventionally real things are. Then when 'There is a pot on the table' is conventionally true, there will be some ultimately true sentence that says there are atoms arranged in such-and-such a way. There is one important caveat here, though: no statement employing *convenient designators* (words for partite entities, such as 'pot') is ultimately truth-apt. In effect this bans all talk of conceptual fictions from our ultimate discourse. This is because if such talk were permitted then we could raise the question whether the pot is identical with or distinct from the atoms arranged pot-wise, and such talk would lead to paradox in short order.⁷ Now it is an open question whether Buddhist Reductionists would want to similarly ban talk of ultimately real entities from the conventional discourse; the texts are silent on this matter. To do so (to make the semantic barrier between the two truths two-way) would mean, for instance, that *sorites* paradoxes could not arise in conventional truth: since one could not then speak of the pot and its constituent atoms in the same breath, one could not ask the embarrassing question at what precise point the pot ceases to exist as one removes randomly selected atoms. Since Buddhist Reductionists think conventional truth is deceptive in important respects, they might see no reason to install such a barrier at the conventional level. Still the conventional truth (i.e. those of our going theories

⁷ One can, however, say things about the word 'pot' at the ultimate level of truth, for instance that the word is a convenient designator (and hence does not designate anything real). This is how Ābhidharmikas avoid paradox while asserting what we might more incautiously put as 'Pots are not ultimately real'. One can keep the ultimate truth contradiction-free by using this sort of meta-linguistic strategy.

that involve commitment to wholes) is useful, and this is difficult to explain if it contains paradoxes. Moreover, there are advantages to be realized if we limit conventional discourse to just talk of conceptual fictions. We can then see how novel predicates can arise at the conventional level for which no algorithm can be given for translation into ultimate discourse, without inviting the mystery-mongering of emergentism. For we can then see how the aesthetic properties of a pot might supervene on the properties of the constituent atoms even though we cannot give a precise formula for arranging atoms so as to produce beautiful pots.

The real payoff from this semantic dualism comes with its application to the concept of the person. We take ourselves to be entities that endure at least a lifetime. Ultimately there are only impersonal impermanent entities making up distinct causally connected series. None of these entities lasts a lifetime. And the series that they constitute, being wholes made of parts, are not ultimately real. But given the facts about such entities and their causal interactions, it turns out to be useful to carve the world up into aggregates that correspond roughly to distinct causal series of psychophysical elements. This in turn makes it conventionally true that we are persons, things that endure at least a lifetime. The personhood concept is a useful fiction.

Our personhood concept carries with it such important features as interest in one's survival, responsibility for one's past actions, and concern about one's future states. Here is an illustration of how such a concept might be a useful fiction. Persons typically take the stance of appropriation toward the past and future states in the causal series: they take those states to be 'me' or 'mine'. Suppose this disposition is the product of early childhood socialization.⁸ Why might it be useful to socialize the child into taking such a stance? Consider the case of brushing one's teeth, something that children typically do not enjoy. The child can have a motive to brush their teeth only if they identify with those future elements that are spared the occurrence of the pain of tooth decay. Only if the child learns to think of that future dental patient as themselves will they see a reason to do something that brings no present pleasure and some minor present discomfort. To instill the concept of the person in this causal series is a way of equipping it with resources that turn it into a center for local enhancement of overall welfare. There is less overall pain and suffering and more overall pleasure and happiness in the world when the child learns to think of itself as a person.

Buddhist Reductionists deny that ultimately there is anything that endures in the way in which we take ourselves to endure. This is sometimes thought to have the consequence that we are not enduring entities. And this in turn would mean that we can have no reason to feel concern about the future states in this causal series. But Buddhist Reductionists call this view about the future 'annihilationism', and

⁸ Buddhist Reductionists typically say that the disposition to engage in appropriation stems from habits acquired in the beginningless series of past lives. We say more about this below.

they consider it one of the two extreme views about ourselves that are to be avoided (the other being the non-Reductionist view that we are separately existing, enduring selves, a view they call 'eternalism'). They advocate what they call a 'middle path' between these two extreme views. Semantic dualism is what makes such a middle path possible. The question 'Am I something that endures?' concerns the temporal properties of a complex entity. Among the presently occurring psychophysical elements, there is no individual element that could take up the first-person stance on its own. The referent of this 'I' could only be something complex. Consequently the question is semantically ill-formed at the ultimate level. No answer to this question could be ultimately true, or ultimately false either. For it concerns something partite, and partite entities can only be referred to at the conventional level. The question only makes sense at that level. And given the facts about the causal series of psychophysical elements, the answer that is conventionally true is that I am something that endures. So the personhood feature of concern about one's future states holds in the only way that it could—conventionally. Likewise for the other important personhood features.

Moreover, if the semantic barrier between the two truths is two-way, there cannot be any incompatibility between the ultimate truth concerning the causal series and the conventional truth about us. Suppose, for instance, that all occurrences of psychophysical elements obey deterministic causal laws. It is then ultimately true that all the elements are causally determined. Might this be incompatible with the statement that persons bear moral responsibility for their actions? For two statements to be incompatible, the truth of one must entail the falsity of the other. And for the entailment relation to hold, the statements in question must be such as to be either true or false in the same way. 'All the elements are causally determined' is, we are supposing, ultimately true. 'Persons bear moral responsibility for their actions' is (for reasons given in the last paragraph) conventionally true. On the hypothesis that the semantic barrier is two-way, neither statement can be true or false in the way that the other is true. So the two statements cannot be incompatible. The claim that determinism is incompatible with the sort of freedom required for moral responsibility cannot be true.⁹

Buddhist Reductionists hold it to be conventionally true that we are persons, things that endure, have concern about their future states, and have the other personhood features. This is conventionally true because things turn out better—there is more overall pleasure and happiness and less overall pain and suffering—when the personhood theory is adopted and employed. But this holds only up to a point. For when we come to think of ourselves as persons, this sets the stage for existential suffering. It gives rise to the feelings of alienation, frustration, and despair that inevitably arise when we confront the fact of our own impermanence. It would be better if we could

⁹ For a more complete discussion of this see Siderits 2008. For a defense of the alternative view that Buddhist Reductionists are committed to hard determinism see Goodman 2002.

reap the benefits of thinking of ourselves as persons—by, for instance, avoiding actions that yield minor present pleasure in exchange for greater long-term pain—while recognizing this for just the useful fiction that it is. The ideal state, they hold, is one in which we continue to behave like persons but cease taking ourselves quite so seriously. This is how, in their view, overcoming our ignorance about our identity can lead to liberation from suffering.

Not all schools of Indian Buddhism espoused Buddhist Reductionism. One clear exception is the school (or group of schools) known as the Personalists (*Pudgala-vādins*).¹⁰ While Personalists agree that there is no self, they hold that the person—understood as a whole that is conceived in dependence on the psychophysical elements—is no mere conceptual fiction. It is not entirely clear just what ontological status they wish to assign it, but they are committed to the claim that the person is neither identical with nor distinct from the elements, as well as that it somehow exists separate from the elements. This suggests a kind of emergentism, the view that in certain cases something genuinely new emerges when sufficiently many entities enter into sufficiently complex interrelations. The most common form of emergentism is property-emergentism, which has it that what emerges is a novel property of the system composed of the more basic entities. The vitalist view that the properties of living systems emerge from the assembly of inanimate material particles is an example of this sort of emergentism.¹¹ But one might also espouse a substance-emergentism, according to which what emerges is a new kind of substance altogether.¹² And this would seem to be what the Personalists had in mind. They agree with Reductionists that the person is conceptualized in dependence on the psychophysical elements, and that the person cannot be said to be either identical with or distinct from the elements. But they reject the claim that the person is thought to exist only because of the useful role the personhood concept plays in our going theory. Their alternative view might be couched in the current parlance as the claim that the person non-reductively supervenes on the psychophysical elements.¹³

¹⁰ For a thorough discussion of the school, its literature, and the difficulties in its interpretation, see the first seven chapters of Priestley 1999. (It should be noted that the interpretation Priestley develops beginning in the eighth chapter is highly speculative.)

¹¹ There is likewise a dispute among Buddhists as to whether there exists anything like a vital force in sentient beings. But this concerned the question whether it is some single ultimately real entity that explains the continued life of a sentient being, or instead that the 'vital force' is a mere conceptual construction reductively supervening on the powers of the being's material components. Neither view represents a kind of property-emergentism. For this dispute see Cox 1995: 291–2.

¹² For a recent formulation and defense of substance emergentism with respect to persons see Hasker 1999.

¹³ For an account of the dispute over non-reductive supervenience in the philosophy of mind see Kim 1998. Buddhist Reductionists claim that the person reductively supervenes on the elements. Their disagreement with the Personalists has to do with whether or not there remains any explanatory work to be done once all the facts about the elements and their arrangement are taken into account. The Personalist view would thus count as an instance of what Parfit calls the 'further-fact' form of Non-Reductionism. See Parfit 1984: 210.

Personalists give several reasons for their claim about persons. Some of these rely on appeals to the Buddha's authority. But Personalists also argue that theirs is the only view that can account for the occurrence of such moral properties as desert. They agree with Buddhist Reductionists that there is no self, so an enduring self cannot be what bears the property of deserving to be rewarded or punished for one's past deeds. But, they claim, if the person were a mere conceptual fiction then it could not do the job either. Buddhist Reductionists obviously disagree, and from what was said earlier we can see how they respond to this objection. But perhaps the greatest challenge for Personalism lies in explaining just how the non-reductive supervenience they claim holds between person and elements is supposed to work. The most sustained attempt to do this comes with their fire-fuel analogy. The idea seems to be that by 'fuel' is meant just those material elements that make up wood and the like, and that fire occurs when certain of those elements are arranged in a certain way. In this case we can say that since fire is one while the elements of which it is composed are many, fire can be neither identical with nor distinct from those elements. But something else is needed if fire is not to be reducible to the elements. Fire must be shown to exert 'downward' causation: it must be the case that the occurrence of fire brings about changes among the elements that cannot be explained on the basis of facts about the elements alone. And this turns out to be singularly difficult to show.¹⁴ Failing such a demonstration, Ockham's razor (known to Indian philosophers as the principle of 'lightness') rules in favor of reducing fire to the elements. For if all that happens at the base level can be explained in terms of facts about the entities at that level, there is no reason to suppose that higher-level entities exist apart from our ways of conceptualizing aggregates of base-level entities. So it is not clear that Personalism constitutes a viable alternative to Buddhist Reductionism.

It is sometimes claimed that, in addition to the Personalists, the Prāsaṅgika Madhyamaka school of Indian Buddhist philosophy also rejects Buddhist Reductionism.¹⁵ This seems plausible given that the simples that Buddhist Reductionism requires for its ultimate ontology are said to be things with intrinsic natures (*svabhāva*), and the central claim of Madhyamaka is that nothing has intrinsic nature. Since Madhyamaka thus rejects the kind of privileged ontology required for semantic dualism, it seems reasonable to suppose that it repudiates the reductive analysis of persons into impersonal elements. But this conclusion may be premature, for it is difficult to reconcile with what Candrakīrti says in commenting on Nāgārjuna's discussion of non-self in *Mūlamadhyamakakārikā* XVIII. 1–5 (Katsura and Siderits 2008: 32–4) and *Mūlamadhyamakakārikā* XXVII. 3–20 (Katsura and Siderits forthcoming). The arguments there are just the ones that Buddhist

¹⁴ The example is discussed at *Abhidharmakośabhāṣya*, 462, translated in Duerlinger 2003: 74–6.

¹⁵ Each of the following makes this claim about Candrakīrti in a slightly different way: Duerlinger 1993; Ganeri 2007; Perrett 2002.

Reductionists typically give for their view. Moreover, if Candrakīrti held a Minimalist view of persons, as Perrett (2002: 378) claims, he should acquiesce in our common-sense concept of persons and the normative commitments it brings with it. Instead Candrakīrti famously likens our ordinary sense of 'I' to having a nest of snakes in the walls of one's house, and he calls realization of non-self the 'magic charm' that destroys all sense of 'I'.¹⁶ These are things a Reductionist might say, but not a Minimalist.

What may have led some to view Candrakīrti as anti-Reductionist is the way he argues for the claim that the person can be neither identical with nor distinct from the elements. Arguing against the view that 'I' denotes all the psychophysical elements in the series (so that the person is identical with the elements) he points out that on one occasion 'I' will denote one subset of those elements, while on another occasion it denotes a distinct subset. There being no single set of elements that my use of 'I' always denotes, the person cannot be identical with any such set. Now this can be taken as an anti-Reductionist argument only if one takes reduction to require full translatability between the two discourses. But Buddhist Reductionism, with its semantic barrier between conventional and ultimate truth, rejects this requirement. This is the point of the commonly cited example of 'The tree bears fruit'. On any given occasion what makes a sentence using the word 'tree' true will be some set of parts in the causal series of elements that constitute an arboreal life history. And among those elements are fruits (assuming for the moment that fruits are ultimate parts). Still they cannot be among the elements involved in assessing the contribution of 'tree' to the sentence 'The tree bears fruit', lest the sentence turn out to be uninformative. Indeed we can appreciate why it would be important that there not be a precise mapping of convenient designators onto ultimately real entities. Such convenient designators as 'tree' and 'pot' are convenient precisely because they are highly tolerant and flexible: they have, as it were, a large margin of error. Given that we have an interest in referring to aggregates with functions like producing fruit and holding honey, and that there are hugely many possible aggregations of particles that can perform these functions, and that our sense organs are macroscopes and not microscopes, and that our computational powers are rather limited, it is in our interest that 'tree' and 'pot' not be identified with any determinate set of particles in any determinate arrangement. This is why Buddhist Reductionism puts semantic barriers between the two truths.¹⁷

¹⁶ See *Madhyamakāvatāra* 6. 141 for the snake analogy, *Madhyamakāvatāra* 6. 145 for the claim about the magic charm. For an English translation see Padmakara Translation Group 2002: 88.

¹⁷ Various logical and semantic strategies have been proposed to circumvent the *sorites* difficulties that would result if terms from the ultimate discourse could be introduced into the conventional discourse (i.e. if the semantic barrier were only one-way). For a brief survey and general discussion see Keefe and Smith 1997: 52–7; also Parsons 2000. In Siderits 2003: 77–85 I suggested that Buddhist Reductionism might be developed in such a way as to yield a semantics for vague terms that employed a degree-theoretic truth-predicate. While I still believe such an approach can be made to work, I am

The example of 'The tree bears fruit' also illustrates a way in which Buddhist Reductionism might be used to explain why we believe that persons have selves. The case of the tree involves taking the set of all the parts in the series and dividing it into two distinct subsets, one taking subject role and the other taking object role. We do something similar with persons, particularly with respect to actions, as in 'I raised my right hand'. This can give rise to the sense that there is some core set of parts that constitutes the control center. But a part that goes in the object slot on one occasion might fall on the subject side on another, as in 'I steered the car around the pylon', where my right hand seems to count as among the entities that are involved in the exercise of control. And we seem capable of exercising some degree of control over all the parts (not on any one occasion, of course, but on successive occasions). This can give rise to the illusion that the control center must be something existing separately from the empirically accessible parts. It can lead to a sense of the self as a 'center of narrative gravity'.¹⁸ It is an illusion, this sense of a transcendent yet contentful self, yet we can see how it would be useful and so would come to play an important role in the common-sense core of our going theory.

Among the many objections that have been raised to Reductionism, those that charge it with circularity are especially important. One such objection is that there cannot be conventional truth without persons to establish the conventions. Since Buddhist Reductionism claims that persons are constituted by the conventions that make up our going theory, it is circular. This objection is easily answered through appeal to selectionist explanations. One can suppose that processes of natural selection yielded aggregate systems of the sort that make up what we are here calling causal series of psychophysical elements. And then, one can suppose, processes of cultural selection yielded groups of such series that shared the conceptual practices making up the personhood theory.¹⁹ To say that it is conventionally true that we are persons is to say that the personhood theory better serves our interests than any of the other readily available theories of what we are. In that case, those social groups in which the theory arose would have outperformed those in which some other conceptual practice arose, leading to the present-day universal acceptance of the personhood theory.

A rather more interesting answer to this objection comes from the Buddhist Reductionist. Buddhists claim that *saṃsāra* is beginningless: there is no first life in the series of lives that has this life as its latest member. (That there may be an end to

now inclined to think that Buddhist Reductionists would reject the modifications of classical logic that introduction of such a truth-predicate requires. This is why I now think they would insist that the semantic barrier be two-way.

¹⁸ See Dennett 1992. See also Flanagan 2002: 213–64. Note though that Flanagan appears to endorse a kind of non-reductive supervenience view of the narrative self (217).

¹⁹ Of course the development of culturally encoded systems of representation requires that there already be proto-conceptual resources in place at the level of the individual. For an example of such bootstrapping see Carey 2004.

this series is the promise of the Buddha's third noble truth, that there is *nirvāṇa* or the cessation of suffering.) If there is no first life in any series of lives, then there need not have been a time before which there was no social structure capable of bringing about socialization into the theory that we are persons. Of course it will be said that the regress cannot be benign, for the time required runs out at the Big Bang. As long as each life lasts for some finite amount of time, there can only be a finite number of lives in a series that began sometime after the Big Bang. But classical Indian cosmologies typically see this universe as just the latest in a beginningless series, the collapse of each sowing the seeds for the start of another. On this view, the Big Bang that started our universe was the effect of a Big Crunch with which the immediately preceding universe ended. One might be highly skeptical about the very hypothesis that there is rebirth. But it is not obvious that this answer to the circularity objection is incoherent.

A second sort of circularity objection is not so easily answered. This can be put as the point that successful reduction requires the existence of experiencing subjects, and Reductionism implicitly calls the existence of such subjects into question. In general, reductionisms seek to show how, given what does exist, it would seem to us as if there is something extra. To reduce heat to mean kinetic energy it is not enough just to show a correlation between judgments of heat and high mean kinetic energy. We need to be convinced both that there is something there for the high mean kinetic energy to be correlated with, and that this something is somehow less than robustly real. The target of the reduction, we say, *just is* certain 'more basic' things. We pull off this trick of making the reduction target something real, but not too real, by making it subjective, a mere appearance, but reliably so. The tangible property of heat is how high mean kinetic energy will seem to systems with sense organs and interests of a certain sort. We can understand how, given our needs and the nature of our cognitive capacities, it would seem to us that there are pots when there are really just atoms arranged pot-wise. But, the objection goes, is it really coherent to assert, as the Reductionist does, that it seems to us as though we are persons when there are really just psychophysical elements arranged person-wise? For then to whom do these appearances appear?

This, it should be noted, is not your garden-variety circularity objection. The complaint is not that the *explanandum* occurs in the *explanans*. That sort of complaint is often lodged against Reductionism, since Reductionist explanations sometimes appeal to the interests served by the personhood concept, and interests would seem to be adjectival on persons. To answer that complaint the Reductionist must show how their explanations can be de-homuncularized, and selectionist stories are one way to do this. But the present objection is different. Its target is not the explanation but the consumer of the explanation. To find a Reductionist explanation plausible I must see my way clear to taking a third-person stance on my own first-person stance. And doing so seems to require my taking a first-person

stance. How, the opponent asks, can it seem to me as if I am a person if there is no 'me' for whom things seem this way?

While those classical Indian philosophers who were realists about the self raised a variety of objections to Buddhist Reductionism, it is interesting that this circularity objection was not among them. The reason may be that there was something in the ultimate ontology of Buddhist Reductionism that could be proposed as the subject of experience without being taken for a self. This is consciousness (*viññāna*), the fifth of the five types of psychophysical element posited by the Buddha. Consciousness is described as that which is aware in general;²⁰ it is another type of psychophysical element, perception (*saṃjñā*), that is said to perform the task of identifying the object of awareness. This makes consciousness look rather like a subject of experience. Great stress is placed, however, on the point that consciousness is impermanent. For instance, the Buddha argues that consciousness originates in dependence on sense-object contact, so that experience across different sense modalities involves distinct consciousnesses.²¹ Since it is one consciousness that serves as the subject of the experience of seeing the mango and another consciousness that serves as the subject of the experience of tasting the mango, consciousness cannot be thought of as a self that serves to unify these distinct experiences. There is, Buddhist Reductionists claim, nothing that unifies these experiences. The appearance of unification arises because of the causal connections that obtain among these and other psychophysical elements in a causal series, plus our use of the convenient designator 'person'.²² So the objection that Reductionism requires an experiencing subject is answered with the claim that the experiencing subject, being momentary, is thoroughly impersonal in nature.

But not all Buddhist Reductionists were happy with this answer. Early in the development of the Yogācāra school of Mahāyāna Buddhism we find a concern to banish all traces of the subject-object dichotomy, including the distinction between conscious subject and object of experience within an individual experience-event.²³ This concern eventually manifests itself in the claim of Dignāga, founder of the school of Buddhist epistemology, that a cognition has two forms, that of the object and that of itself as something that cognizes. This claim is directed against the naive realist view of cognition, according to which cognitions only have the single nature of cognizing, the form of the object being located in an object that is wholly distinct from that which cognizes. Dignāga is here

²⁰ See, for instance, the definition of *viññāna* (consciousness) given at *Milindapāṇa*, 62, tr. at Rhys Davids 1965: 95.

²¹ See, for instance, Horner 1954: 314–15.

²² This is the subject of the second half of Vasubandhu's 'Refutation of the Theory of the Person', the ninth chapter of *Abhidharmakośabhāṣya* (the first half being taken up with the investigation of Personalism). For translations and discussion see Duerlinger 2003; Kapstein 2001: 367–75.

²³ See Vasubandhu's commentary on *Vimsatikā*, 10, Sthiramati's commentary on *Trimśikā*, 21cd in Levi (1925).

representing the view of the representationalist and the subjective idealist, for both of whom what one is directly aware of in cognition is in some sense immanent to the cognizer. But it is open to the representationalist and the idealist to hold that the form of the object occurs in a mental image, and that awareness of that form occurs through the relation of the image to a distinct consciousness event that cognizes the image. Dignāga's claim that both forms are in a single cognition reflects the Yogācāra concern that any type of subject–object dichotomy, even one based on a distinction among momentary mental events, can serve as a breeding ground for a sense of 'I' and thus lead to suffering. This claim would give the Buddhist Reductionist one way to answer the present circularity objection, which in effect charges the Reductionist with trying to reduce everything to the third-person stance of objectivity through appeal to the subjective and the first-person perspective. Dignāga would reply that the subject–object dichotomy is merely a useful conceptual construction, one that a proper (Reductionist) understanding of the nature of cognition helps us transcend.

One might wonder whether this reply can be made to work. But perhaps a more important question here is what reason there is to hold that cognitions have two forms. Dignāga argues that if a cognition only had one form, either the form of the object or the form of itself as something that cognizes, then we would be unable to distinguish between the cognition of a pot (c_1) and the cognition of the cognition of a pot (c_2). For if a cognition is characterized only by the form of its object, then c_1 and c_2 would both have the nature of the form of a pot, P; while if a cognition is only characterized by its own form C, then once again c_1 and c_2 would share the same form. But this difficulty only arises if we assume that what is cognized by a cognition must be immanent to that very cognition. This assumption is rejected by direct realists, who claim that a cognition always cognizes something distinct from itself. They argue that to say otherwise is to hold that a cognition can cognize itself, something we know to be no more possible than that an acrobat can stand on their own shoulders. One can then distinguish between c_1 and c_2 on the basis of the fact that while c_1 and c_2 both have C as their nature, c_1 has something P-natured as its object while c_2 has something C-natured as its object.

Dignāga would reply by defending the claim that cognition cognizes itself, which is supported by the analogy of light that illuminates both the object and itself. He also argues that the alternative view, that consciousness is irreflexive, suffers from two major difficulties. First, it leads to an infinite regress: if a second cognition is required in order to cognize my cognition of a pot, then a third is required to cognize the second, etc. Second, the notion that cognition is itself cognized by a distinct act of introspection is incoherent. Presumably the introspective cognition occurs after the cognition that it takes as object. But this turns introspection into a kind of remembering: to be aware of the cognition of a pot is to remember cognizing the pot a moment earlier. And, points out Dignāga, one cannot remember what one did not oneself experience. So I cannot introspectively cognize an

earlier cognition if that cognition was not itself something I was aware of at the time it occurred. Given that we can be aware of our own cognizing, it must be the case that cognition cognizes itself as well as the form of the object. So the argument for cognition's having two forms goes through.

This defense can be questioned. First, the analogy of the light that illuminates itself is problematic. True, when we shine a torch we may say that we see not just the objects in the room but the light as well, and this without some second light that illuminates the first. But when we see a beam of light it is actually illuminated dust motes that we see. Second, the infinite regress argument presupposes what it tries to prove. Just as other-illumination (the view that a cognition is only illuminated by another cognition) means that cognition of c_1 requires a subsequent c_2 , so the cognition of c_2 would require yet another cognition c_3 . But why should this mean that on that view the cognition of a pot would require an infinite number of cognitions? On the other-illumination view, apperception is always optional: one can cognize a pot without cognizing the cognition that cognizes the pot. The alleged infinite regress arises only if one assumes that a cognition cannot cognize its object unless it cognizes itself. Third, the argument from memory is effective only against that form of other-illumination view that sees apperception as introspection. According to the form of other-illumination view espoused by the (Brahmanical) Bhāṭṭa Mīmāṃsā school, a cognition is cognized only through a kind of abductive inference. This view is formulated based on the assumption that a cognition can never be directly cognized. Our awareness of our cognition comes about instead through an inference to the best explanation: the pot now having the property of cognizedness, there must have occurred a cognition taking the pot as its object. It thereby avoids the objection from memory by in effect accepting the claim that cognition could be directly cognized only if it were reflexive, but denying that cognition is directly cognized. Finally, questions can be raised about the metaphysics of the claim that cognition has two forms, which comes perilously close to the substantialist view all Buddhist Reductionists reject.

Given these difficulties, it might be worth seeing what other options are open to the Buddhist Reductionist. Not all Buddhists accepted Dignāga's claim that consciousness is reflexively aware; some held the other-illumination view. But I know of no surviving text that defends this view and discusses how it could be used to answer the circularity objection we have been considering. So what follows must be somewhat speculative. Suppose we wished to say that our awareness of our own cognition comes about through a kind of abductive inference. We would then be claiming that awareness is something I posit in order to explain the cognizedness I experience in the pot and the like. But now the question arises how one can be aware of cognizedness in the object without being aware of cognition. Cognizedness looks like a relational property, and one can be aware of relational properties only if one can be aware of the relata: one can be aware of Jill's property of being taller than Jack only if one can also be aware of Jack and his height. To this

objection some Bhāṭṭas reply that the property of cognizedness is a property the pot has just by virtue of its being available for action and speech.²⁴ The difficulty with this response is that it runs the risk of turning cognition into a mere empty place-holder. To see why, consider the case of blind-sight, where the patient successfully navigates around all the obstacles in the path while sincerely claiming to see nothing. We think of this as a case of 'unconscious seeing' precisely because visual input is going directly to the action-control system without being made available to other systems like speech and memory. This suggests that for the normally sighted subject the obstacle has the property of cognizedness just by virtue of visual input being available to several different systems. And if the distinct property of cognizedness is not found in blind-sight, where there is just the one input-output path, why suppose that such a distinct property emerges where there are several paths possible? To a Buddhist Reductionist this will look like a case of reifying what is merely a convenient designator. The upshot is that if the other-illumination theorist goes the abductive inference route to account for our awareness of our own awareness, consciousness turns into a conceptual construction and not something ultimately real.

Buddhist Reductionists would be loath to accept this consequence. First, the Buddha seems to have considered consciousness to be a psychophysical element, so calling it a conceptual construction is tantamount to calling the Buddha's authority into question. Second, the practice of meditation looks like nothing so much as the disciplined investigation of consciousness, and meditation is supposedly efficacious in producing liberation, so this might raise doubts about the Buddhist path to *nirvāṇa*. Still I think the other-illumination approach might help the Buddhist Reductionist answer the circularity objection. The first point to make here is that to take this approach is not necessarily to embrace eliminativism about consciousness. The claim is that consciousness, while not part of our final ontology, is posited by a theory we use in coping with reality. Both a reductionist and an eliminativist would accept this claim; their disagreement concerns whether the theory is useful or useless. Just as Buddhist Reductionists claim the theory that we are persons is useful, so they could claim a similar status for the theory that posits consciousness as a psychophysical element. Buddhist Reductionists could say that consciousness is conventionally real.

More importantly, this strategy would give the Buddhist Reductionist a response to the circularity objection that does not rely on the notion of reflexive awareness. The objection was that Reductionism cannot remove the source of the sense of 'I am' in the guise of the experiencing subject (even if that subject is only momentary). The response is that consciousness as cognizer is constructed in the interest of achieving conceptual economy. There are just the causal paths linking

²⁴ There is an interesting parallel here with the notion of global availability discussed in Metzinger 2003: 31.

input and output; consciousness is what we call the situation of there being a multiplicity of open paths. Of course it will be objected that we can never be persuaded that this is true, since being persuaded would require an occurrence of the what-it-is-like-ness characteristic of the (conscious) state of conviction. To this, two responses are possible. The first is that some illusions are not dispelled by knowledge of their sources. One can know perfectly well that the lines in a demonstration of the Ponzo illusion are of equal length, yet one will continue to see some as longer than others. The sense that there is the subjective or perspectival, it might be said, is just such an illusion. Of course this must count as the mother of all illusions, since without it there could be no gap between how things seem and how things are. But it nonetheless makes sense to speak of it as an illusion, since it leads the system in which it occurs to misrepresent its environment. There is no what-it-is-like-ness for the thermostat, yet we speak naturally of our fooling the thermostat into taking the room to be warmer than it is by holding a candle nearby. So we might say that the psychophysical system misrepresents its environment as containing a subject to which things seem a certain way. One's grasping of this point may include an instance of just such a seeming. But this need not be paradoxical. Once we see how the illusion is generated, and why it persists (due to its utility for systems like these), we can understand that it is a misrepresentation, and why such a misrepresentation occurs.

A second possible response is that this is among those illusions that are dispelled by knowledge of their sources. In certain Buddhist accounts of the fully enlightened being we glimpse the possibility of what might be called a state of total blind-sight. According to the teachings of Mahāyāna Buddhism, the fully enlightened being works ceaselessly to help the unenlightened overcome suffering. On some accounts the enlightened being does this without cognizing those being helped. We might dismiss accounts of such a 'Robo-Buddha' as no more than devotionalist hyperbole. But another possibility is suggested by a perfectly sane piece of Buddhist epistemological doctrine: the claim that while reliably caused perception in familiar circumstances generates knowledge immediately, reliably caused perception in a novel setting results in knowledge only after subsequent confirmation. The idea is that in the familiar setting there is no need to run the fact-checker routine, one simply acts in accordance with the input and one's interests, while in unfamiliar circumstances there is a need to reflect on the sources of the input. Such reflection takes the form of treating the input as an internal state of the system, something made available to multiple sub-systems of the organism. So the suggestion is that the positing of cognition as an inner state occurs when we don't know our way about. Robo-Buddhas, having developed their skills at exercising compassion through countless eons of training, do know their way about. I am not such a being. But being thirsty, and having quenched my thirst hundreds of times at the water fountain in front of me, I act effortlessly on my present visual stimulation. There is no need to treat my visual stimulation as an inner cognitive state. Were I asked if

I was aware of seeing the fountain I would doubtless say yes. But this judgment is retrospective. The suggestion is that I say this only because often I do not know my way about, and so have to check up on the credentials of my sensory stimulations. Robo-Buddhas, being fully enlightened and so always knowing their way about, are able to act directly on their sensory stimulation without making it available to other systems. Having total blind-sight, for them the illusion of consciousness is dispelled. There being nothing for which things appear a certain way, there is no longer the basis for a sense of 'I am'. There is just a causal series of psychophysical elements, interacting with its environment in such a way as to maximize overall welfare.

REFERENCES

- ALBAHARI, MIRI (2007). *Analytical Buddhism: The Two-Tiered Illusion of the Self* (New York: Palgrave Macmillan).
- CAREY, SUSAN (2004). 'Bootstrapping and the Origins of Concepts', *Daedalus*, 59–68.
- COX, COLLETT (1995). *Disputed Dharmas* (Tokyo: International Institute for Buddhist Studies).
- DENNETT, DANIEL C. (1992). 'The Self as a Center of Narrative Gravity': <http://ase.tufts.edu/cogstud/papers/selfctr.htm> (accessed July 2010).
- DUERLINGER, JAMES (1993). 'Reductionist and Non-Reductionist Theories of Persons in Indian Buddhist Philosophy', *Journal of Indian Philosophy*, 21: 79–101.
- (2003). *Indian Buddhist Theories of Persons: Vasubandhu's 'Refutation of the Theory of a Self'* (London: RoutledgeCurzon).
- FLANAGAN, OWEN (2002). *The Problem of the Soul* (New York: Basic Books).
- GANERI, JONARDON (2007). *The Concealed Art of the Soul* (New York: Oxford University Press).
- GOODMAN, CHARLES (2002). 'Resentment and Reality', *American Philosophical Quarterly*, 39: 359–72.
- HASKER, WILLIAM (1999). *The Emergent Self* (Ithaca, NY: Cornell University Press).
- HORNER, L. B. (tr.) (1954). *The Collection of Middle Length Sayings* (Majjhima-Nikaya) (London: Pali Text Society).
- KAPSTEIN, MATTHEW (2001). *Reason's Traces: Identity and Interpretation in Indian and Tibetan Buddhist Thought* (Studies in Indian and Tibetan Buddhism; Boston: Wisdom).
- KATSURA SHORYU and SIDERITS, MARK (tr.) (2008). *Mūlamadhyamakakārikā XI–XXI, Journal of Indian and Tibetan Studies (Indogakuchibettogaku Kenkyū)*, 12: 170–221.
- and — (tr.) (forthcoming), *Mūlamadhyamakakārikā XXII–XXVII*.
- KEEFE, ROSANNA, and SMITH, PETER (eds) (1997). *Vagueness: A Reader* (Cambridge, Mass.: MIT Press).
- KIM, JAEGWON (1998). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation* (Cambridge, Mass.: MIT Press).
- LEVI, SYLVAIN (ed.) (1925). *Vijnaptimatratasiddhi: Deux traites de Vasubandhu: Vimsatika (La vingtaine) accompagnée d'une explication en prose, et Trimsika (La trentaine) avec le commentaire de Sthiramati* (Paris: Champion).

- MERRICKS, TRENTON (2001). *Objects and Persons* (Oxford: Clarendon Press).
- METZINGER, THOMAS (2003). *Being No One* (Cambridge, Mass.: MIT).
- Padmakara Translation Group (2002). *Introduction to the Middle Way* (Boston: Shambala).
- PARFIT, DEREK (1984). *Reasons and Persons* (Oxford: Clarendon Press).
- PARSONS, TERRENCE (2000). *Indeterminate Identity* (Oxford: Clarendon Press).
- PERRETT, ROY (2002). 'Personal Identity, Minimalism and Madhyamaka', *Philosophy East and West*, 52: 373–85.
- PRIESTLEY, LEONARD (1999). *Pudgalavada Buddhism: The Reality of the Indeterminate Self* (South Asian Studies Papers, 12; monograph 1; Toronto: Centre for South Asian Studies).
- RHYS DAVIDS, T. W. (tr.) (1965). *The Questions of King Milinda* (Delhi: Motilal Banarsidass).
- SIDERITS, MARK (2003). *Personal Identity and Buddhist Philosophy: Empty Persons* (Farnham: Ashgate).
- (2008). 'Paleo-Compatibilism and Buddhist Reductionism', *Sophia*, 47: 29–42.
- VAN INWAGEN, PETER (1995). *Material Beings* (Ithaca, NY: Cornell University Press).

THE OXFORD HANDBOOK OF

THE SELF

Edited by

SHAUN GALLAGHER

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford ox2 6dp

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi

Kuala Lumpur Madrid Melbourne Mexico City Nairobi

New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece

Guatemala Hungary Italy Japan Poland Portugal Singapore

South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States

by Oxford University Press Inc., New York

© The several contributors 2011

The moral rights of the authors have been asserted

Database right Oxford University Press (maker)

First published 2011

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
without the prior permission in writing of Oxford University Press,
or as expressly permitted by law, or under terms agreed with the appropriate
reprographics rights organization. Enquiries concerning reproduction
outside the scope of the above should be sent to the Rights Department,
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover
and you must impose the same condition on any acquirer

British Library Cataloguing in Publication Data

Data available

Library of Congress Cataloging in Publication Data

Data available

Typeset by SPI Publisher Services, Pondicherry, India

Printed in Great Britain

on acid-free paper by

MPG Books Group, Bodmin and King's Lynn

ISBN 978-0-19-954801-9

3 5 7 9 10 8 6 4 2