

Enacting the self: Buddhist and Enactivist approaches to the emergence of the self

Matthew MacKenzie

Published online: 23 May 2009
© Springer Science + Business Media B.V. 2009

Abstract In this paper, I take up the problem of the self through bringing together the insights, while correcting some of the shortcomings, of Indo–Tibetan Buddhist and enactivist accounts of the self. I begin with an examination of the Buddhist theory of non-self (*anātman*) and the rigorously reductionist interpretation of this doctrine developed by the Abhidharma school of Buddhism. After discussing some of the fundamental problems for Buddhist reductionism, I turn to the enactive approach to philosophy of mind and cognitive science. I argue that human beings, as dynamic systems, are characterized by a high degree of self-organizing autonomy. Therefore, human beings are not reducible to the more basic mental and physical events that constitute them. I critically examine Francisco Varela’s enactivist account of the self as virtual and his use of Buddhist ideas in support of this view. I argue, in contrast, that while the self is emergent and constructed, it is not merely virtual. Finally I sketch a Buddhist-enactivist account of the self. I argue for a non-reductionist view of the self as an active, embodied, embedded, self-organizing process—what the Buddhists call ‘I’-making (*ahaṃkāra*). This emergent process of self-making is grounded in the fundamentally recursive processes that characterize lived experience: *autopoiesis* at the biological level, *temporalization* and *self-reference* at the level of conscious experience, and conceptual and narrative *construction* at the level of intersubjectivity. In Buddhist terms, I will develop an account of the self as dependently originated and empty, but nevertheless real.

Keywords Buddhism · Enaction · Self · Reductionism · Emergence

M. MacKenzie (✉)
Colorado State University, 233 Eddy Hall, Fort Collins, CO 80523, USA
e-mail: matt.mackenzie@colostate.edu

Introduction

The conception of the self as a substance separate from the body and the rest of the natural world (e.g., the Cartesian ego) is widely rejected today. Yet many accounts of the self are developed based on assumptions, such as substantialism and objectivism, that arguably remain basically Cartesian (cf. Dennett 1991; Metzinger 2003). In contrast, both Buddhism and enactivism present fruitful alternatives to broadly Cartesian approaches to cognition, subjectivity, embodiment, and the nature of the self. Indeed, the enactive approach to cognition and its allied method of neurophenomenology explicitly and systematically draw from Buddhist thinkers, ideas, and practices in order to move beyond Cartesianism. In this paper, I take up the problem of the self through bringing together the insights, while correcting some of the shortcomings, of Indo-Tibetan Buddhist and enactivist accounts of the self. I begin with an examination of the Buddhist theory of non-self (*anātman*) and the rigorously reductionist interpretation of this doctrine developed by the Abhidharma school of Buddhism. After discussing some of the fundamental problems for Buddhist reductionism, I turn to the enactive approach to philosophy of mind and cognitive science. In particular, I argue that human beings, as dynamic systems, are characterized by a high degree of self-organizing autonomy. Therefore, human beings are not reducible to the more basic mental and physical events that constitute them. In a similar vein, Francisco Varela argues that the self emerges through the processes of self-organization, and that the self is thus merely *virtual* (Varela 1999). I critically examine Varela's enactivist account of the self as virtual and his use of Buddhist ideas in support of this view. I argue, in contrast, that while the self is emergent and constructed, it is not merely virtual. Finally, I sketch a Buddhist-enactivist account of the self. I argue for a non-reductionist¹ view of the self as an active, embodied, embedded, self-organizing process—what the Buddhists call 'I'-making (*ahaṃkāra*). This emergent process of self-making is grounded in the fundamentally recursive processes that characterize lived experience: *autopoiesis* at the biological level, *temporalization* and *self-reference* at the level of conscious experience, and conceptual and narrative *construction* at the level of intersubjectivity. In Buddhist terms, I will develop an account of the self as dependently originated and empty, but nevertheless real.

Non-self

The doctrine of non-self (*anātman*) is perhaps the best known and most controversial aspect of Buddhist thought. On the Buddhist view, phenomena arise in dependence on a network of causes and conditions. This is the fundamental Buddhist notion of interdependent arising (*pratītyasamutpāda*). The Buddhist analysis of any particular entity, event, or process will focus on the dynamic patterns of interaction within and through which it arises, has its effects, and passes away. It is against the backdrop of

¹ 'Reductionism' is often used very liberally in the literature on personal identity, such that an account of personal identity is reductionist so long as it does not rely on either a Cartesian ego or a brute 'further fact'. My view does not easily fit into these categories, but rather is an emergentist self-constitution view of the self.

this basic analytical and ontological commitment that we can understand the Buddhist account of the self.

First and foremost, the doctrine of non-self is a rejection of the *ātman*, the enduring substantial self. On this view, the ‘self’ (*ātman*) is not just another term for the empirical person (*pudgala*), but is rather the substantial, essential core of the person—the inner self whose existence grounds the identity of the person. Within the Brahmanical religious and philosophical tradition, the *ātman* is generally given a strongly metaphysical interpretation. It is the unitary, essentially unchanging, eternal, spiritual substance that is said to be one’s true self. However, the ultimate target of the Buddhist theory of non-self is not the rarified spiritual conception of self commonly defended by various Brahmanical schools. Most fundamentally, the Buddhist target is a much more widely held and more deeply entrenched conception of the self. Galen Strawson’s account of our basic sense of self fits well with Buddhist characterizations of the *ātman*. He writes:

I propose that the mental self is ordinarily conceived or experienced as:

- (1) a *thing*, in some robust sense
- (2) a *mental thing*, in some sense
- (3, 4) a *single thing* that is single both *synchronically* considered and *diachronically* considered
- (5) *ontically distinct* from all other things
- (6) a *subject of experience*, a conscious feeler and thinker
- (7) an *agent*
- (8) a thing that has a certain character or *personality* (Strawson 1999: 3).

Compare Strawson’s view to Miri Albahari’s account of the *ātman* (Pāli: *atta*) in early Buddhism:

A self is defined as a bounded, happiness-seeking/*dukkha* [suffering]-avoiding (witnessing) subject that is a personal owner and controlling agent, and which is unified and unconstructed, with unbroken and invariable presence from one moment to the next, as well as with longer-term endurance and invariability (Albahari 2006: 73).

Here we find the self understood as an experiencing subject, owner, and controller that is bounded (or, in Strawson’s terms, ‘ontically distinct’) and enduring. Albahari also usefully distinguishes between the subject and the self. A subject is “*witnessing* [awareness] *as it presents from a psycho-physical (hence spatio-temporal) perspective*” (Albahari 2006: 8, emphasis in original). Thus, while perspectival experience implies a subject, it does not necessarily imply a self. For a self is a particular *type* of subject: bounded, enduring, controlling, and so on. Central to the Buddhists, the self in the above sense does *not* exist and our deeply entrenched sense that we are such an entity is at the root of our existential and spiritual bondage (*samsāra*).

Rejecting the existence of the substantial self, the Buddhists argue the existence of a person (*pudgala*) consists in the existence of the five *skandhas* (bundles or aggregates) organized in the right way. The five *skandhas* are:

1. *Rūpa*: the body or corporeality
2. *Vedanā*: affect and sensation

3. *Samjñā*: perception and cognition
4. *Samskāra*: conditioning and volition
5. *Vijñāna*: consciousness

These five *skandhas* are not to be taken as independent things, but instead are seen as interdependent aspects of a causally and functionally integrated psycho-physical (*nāma-rūpa*) system or process (*skandhasantāna*: an ‘aggregate-stream’ or ‘bundle-continuum’).

The *rūpa-skandha* (material form) refers to the corporeal aspect of the human being, including the organizational structure of the person as an organism. The *vedanā-skandha* denotes the qualitative and affective dimensions of the person and her experience—that is, both basic sensory qualia (the look of red, a sharp pain, the taste of water) and the affective tone of experience (pleasant, unpleasant, or neutral). The *saṃjñā-skandha* denotes the more fully cognitive faculty of perception, including the ability to identify and re-identify objects of experience.² The operation of this capacity depends on sensory contact (*sparśa*) with the environment as well as sensory-motor skills (such as exploratory behavior) and is often taken to involve the use of concepts. Next, the *saṃskāra-skandha* (conditioning) includes the various dispositions, capacities, and formations—such as sensory-motor skills, memories, habits, emotional dispositions, volitions, and cognitive schemas—that both enable and constrain the person and her experiences. This category also includes our basic conative impulses—attraction, repulsion, and indifference—which are in turn closely tied to our feelings and the affective modalities (*vedanā*) of experience. On the Buddhist view, typically one’s whole being in the world is driven by this sedimented conditioning—and not always for the better. Indeed, the basic conative impulses often manifest in pathological ways, as the ‘three poisons’ of greed, hatred, and ignorance. As Dan Lusthaus remarks, “such predilections are always already inscribed in our flesh, in our very way of being in the world, even while we ignore—or remain ignorant of—the causes and conditions that have given rise to them” (Lusthaus 2002: 49). Finally, the *vijñāna-skandha* denotes discerning or discriminating intentional consciousness that integrates and is conditioned by the other *skandhas*.

Therefore, in the standard Buddhist analysis, the person is not an entity that can exist independently of the five *skandhas*. Take away the complex, impermanent, changing *skandhas* and we are not left with a constant, substantial self; we are left with nothing. Moreover, the diachronic identity of a person consists in the appropriate degree of continuity and connectedness of the *skandhas*—that is, it is a matter of there being a causally and functionally integrated series or stream of *skandhas*.

Having briefly sketched the theory of non-self, let us examine two lines of argument against the existence of the self (*ātman*): the *critical* argument and the *epistemic* argument. First, it is argued that none of the *skandhas* individually, nor the whole complex of *skandhas* could be the self—i.e., the independent, substantial,

² The term ‘*saṃjñā*’ (*sam*: ‘to put together’ + *jñā*: ‘knowledge’) is cognate to ‘cognize’ and can have the sense of ‘synthesis’ as well as ‘association’. Lusthaus translates *saṃjñā* as “associational knowledge” (Lusthaus 2002: 47).

enduring, inner controller and owner of the *skandhas*. Upon examination, none of the five *skandhas* meets these criteria of selfhood.³ The various mental factors (*nāma-skandha*) are simply too transitory, too mutable to constitute the stable, enduring essence of the person. Moreover, the mental factors are revealed in experience as a stream (*santāna*) or flow, rather than as a substance or object. The body is perhaps more stable, but the fundamental problem is the same: like any complex phenomenon, the body is in perpetual flux. How should we specify the persistence-conditions of the body? One might attempt to identify the body's unique ontological boundary or some essential part of the body that explains its persistence. But neither of these strategies looks particularly viable. The physical boundaries of the body are vague and even if one could find the essential part of the body, it is doubtful that this essential part could meet the other criteria of selfhood. Thus it appears that none of the *skandhas* individually, neither mental factors nor the body, could be the substantial enduring self.

What, then, of the *skandhas* taken together, the *nāma-rūpa* or psycho-physical complex? Could this be the self? One problem with this response is that the psycho-physical complex is the empirical person, whereas the self is being posited as the essence of the person which grounds and explains the persistence of the person. The empirical person, like the individual *skandhas*, is in flux and therefore its endurance is equally problematic. Therefore, to simply identify the self with the person as a whole would be to conflate the explanans with the explanandum. Secondly, the relationship between a whole and its parts is problematic and the Buddhists deny that a complex could be the *independent* owner and controller of its parts. Therefore, the substantial, essential self is not found among the *skandhas* individually or collectively.

The second, and later, line of argument builds on the first. According to the Buddhist philosopher Vasubandhu (4th century CE), we must apprehend the self either through direct acquaintance or through inference.⁴ But we do not apprehend the self through either means. Therefore, we have no epistemic warrant for the existence of the *ātman*. The self is not a direct object of the five external senses or introspection. And while human beings typically have a *sense* of self, it does not follow from this that the sense of self provides direct acquaintance with an enduring, substantial self. Moreover, while it is certainly possible for 'the self' to be an object of thought, again it does not follow that the self exists.

So if the self is not known through direct acquaintance, then perhaps it is known through inference. Vasubandhu examines a valid inference to the existence of the unobservable sense faculties and then asserts that there is no such valid inference to the existence of the unobservable self. In the case of the sense faculties, there is some reasonable way to tell whether the sense faculties are present or absent (e.g., in case of the blind person versus the sighted person). Can the same be said for the *ātman*? One might, for example, posit the existence of distinct substantial selves in order to individuate persons A and B. But because the substantial self is supposed to

³ The classic version of the criterial argument occurs in *Samyutta Nikāya* 3.66-68. Cf. Holder (2006) for a translation.

⁴ This argument occurs in the "Refutation of the Theory of the Self" 1.2. Cf. Duerlinger (2003) for a translation.

retain its identity independently of the ever-changing stream of mental and physical events associated with A and B, how are we to establish anything about these posited selves? As it stands, the empirical evidence—e.g., distinct bodies, various uses of names and ‘I’—is consistent with both the presence and the absence of the posited self, as well as a single shared self or a new substantial self each moment. Hence, the inference to the existence of the *ātman* looks weak. Vasubandhu’s assertion here is not decisive, but the underlying argumentative strategy is to shift the burden of proof onto the proponent of the substantial self. Is there inferential warrant for positing an enduring substantial self or can the phenomena (e.g., memory) be accounted for in terms of the systematic relations between various mental and physical events and processes? The Buddhists, of course, opt for the latter approach on grounds of epistemic and ontological parsimony. We are, they argue, *selfless persons* (*pudgalanairātmyā*).

Buddhist reductionism

The account of human beings as selfless persons is held by all major Buddhist schools, but there has been a great deal of disagreement as to the full ontological implications of the rejection of a substantial self. For the Abhidharma or Buddhist reductionist schools, the doctrine of *anātman* is at the center of a radically reductionist, anti-substantialist empiricism. Everyday entities such as pots and people are not ontologically basic (*dravyasat*), but rather are reducible to aggregations of basic entities. On this view, the seemingly objective, mind-independent unity of everyday composite objects is illusory—these entities have only a secondary, conceptual existence (*prajñaptisat*). The ontologically basic entities to which everyday things are reducible are called *dharmas*. These are simple, fleeting events individuated by their intrinsic defining characteristic (*svalakṣaṇa*). Moreover, the Ābhidharmika’s basic ontology is fairly austere—according to one school,⁵ there are only seventy-five types of *dharmas*. As the Abhidharma philosopher Vasubandhu explains the view:

That of which one does not have a cognition when it has been broken is real in a concealing way (*saṃvṛti-sat*); an example is a pot. And that of which one does not have a cognition when other [elemental qualities (*dharma*)] have been excluded from it by the mind is also conventionally real; an example is water. That which is otherwise is ultimately real (*paramārtha-sat*). (Vasubandhu AK 6.4)

This view constitutes a type of anti-realism about everyday composite entities, including persons. Such entities may be pragmatically or conventionally real (*saṃvṛtisat*), but they are not ultimately real (*paramārthasat*). The being of these entities is fully accounted for in terms of more basic entities; they are fully

⁵ The school in question is the Sarvāstivāda Vaibhāṣika. When referring to Abhidharma and Buddhist reductionism, it is this school I have in mind.

analytically and ontologically decomposable. Thus, they have a merely derived nature (*parabhāva*), rather than their own irreducible intrinsic nature (*svabhāva*). Further, conventionally real entities must be epiphenomenal because if they were to have their own causal powers, they would not be completely reducible. Hence, according to the Ābhidharmikas, all causation is microcausation—that is, real causation occurs only between simple, momentary *dharmas*. Further, the genuine causal powers of these entities are determined by their intrinsic natures. Notice, then, that this two-tiered ontology rests on a radical dichotomy between the entities with a purely extrinsic nature (*parabhāva*) and those with a purely intrinsic nature (*svabhāva*).

Given such a revisionist ontology, one can see the importance of the Buddhist doctrine of two truths. On the Ābhidharmika view, conventional truths (*saṃvṛtisatya*) are those truths that quantify over reducible or conventionally real (*saṃvṛtisat*) entities, whereas ultimate truths (*paramārthasatya*) only quantify over irreducible or ultimately real (*paramārthasat*) entities. When using conventional discourse, one is not ontologically committed to anything but the entities mentioned in the ultimate discourse, even if conventional discourse is not analytically reducible to ultimate discourse. Further, the discourse of ultimate truth is the Ābhidharmika's 'philosophically favored discourse'—that is, the discourse in terms of which all other discourses are ultimately to be explained (Arnold 2005).

Persons, then, are organized, temporally extended systems of mental and physical events characterized by dense causal and functional interconnectedness, including complex physical and psychological feedback loops. Psycho-physical systems are also seen as deeply intertwined with and dependent upon the larger environment. Indeed, for the Buddhist reductionist, there is no sharp dividing line between the collection of events labeled "person" and the collection of events labeled "environment." These terms do not carve the world at its joints; they are pragmatic, interest-relative categories. In order to understand the psychological dynamics that give rise to and perpetuate suffering, one does not look for a substantial mental self or an enduring substantial person. Instead, the Buddhist analyst attempts to understand the complex interrelations between mental and physical events over time. The rigorous Abhidharma analysis, though, goes beyond the early Buddhist shift in perspective from personal to impersonal analysis and defends a strict mereological reductionism—this position, I argue in the next section, is in significant tension with the more general Buddhist analysis of sentient beings.

Four problems for Buddhist reductionism

While Buddhist reductionism offers a powerful critique of and a sophisticated alternative to substantialist views of the self, this radical view of the human person did not go unchallenged. Indeed, both Buddhist and non-Buddhist philosophers vigorously disputed the Abhidharma or Buddhist reductionist approach. In this section, we will examine four interconnected problems for Buddhist reductionism, and for strongly reductionist theories of the person in general. These four problems are: personal and experiential continuity, first-person consciousness, mereological reductionism, and the reification of *dharmas*. I argue that a turn to the enactive

approach will contribute to the development of an anti-substantialist account of the person that overcomes these problems.

The first problem for the Buddhist reductionist has to do with personal continuity and, even more fundamentally, the continuity of the pre-personal body-mind stream that is the ground of personal continuity. An advocate of the self (*ātmavādin*) will want an account of diachronic personal identity (or at least personal continuity⁶) in the absence of a self. And as we have seen, the Buddhist reductionist holds that personal continuity is reducible to psychological continuity (memory, skills, habits, personality traits), which is in turn reducible to causal connections between impermanent mental and physical events. One problem for this approach is that there are just too many causal connections. By their own view, the world is taken to be a causally interdependent network of events. How are we to individuate different streams of events?

The Buddhist reductionist has a two-part response to this problem. At the conventional (*saṃvṛti*) level, streams are individuated by the *density* of causal connections and by the way in which some sets of interconnected mental and physical events are able to ground relatively stable capacities or functions, such as perception and motility.⁷ Furthermore, at the ultimate (*paramārtha*) level, there simply is no ontologically correct way to individuate streams. Individuating streams (carving them out of the causal manifold) is an inherently pragmatic and interest-relative activity, and thus at this level of analysis, there is no fact of the matter about the identity of streams.

However, there is a deeper problem here, pointed out by the *Vaiśeṣika* philosopher, Śrīdhara (c. 990 CE).

[Buddhist:] As a result of there being a causal connection, a later memory [is a memory] of what was experienced at an earlier moment. The son does not, however, remember what was experienced by the father; this is because there is no causal connection between the cognitions of a father and son, and their bodies, though admittedly so [connected], are not [themselves constituted of] consciousness.

[Śrīdhara:] This is not well-reasoned, for in the absence of self, there would be no determinate notion (*niścaya*) of a causal connection. At the time of the cause, the effect has yet to occur, and when its time comes, the cause has gone. Aside from the two of them, some unitary perceiver is denied; so who would observe the causal connection between those two things occurring in sequence? (Ganeri 2007: 177)

In order to form a determinate notion of a causal sequence or stream, Śrīdhara argues, one must be able to *experience* a causal sequence.⁸ And yet, on the radically reductionist view of the Ābhidharmikas, there are only momentary (*kṣanika*) events

⁶ Continuity is a weaker relation than identity in that continuity comes in degrees, while identity is all or nothing. Non-substantialist theories of the person typically account for diachronic personal identity in terms of continuity.

⁷ Though the reductionist still owes an account of how to get from causal connections to the kind of semantic and (broadly) narrative connections that seem to play an important role in any plausible account of psychological continuity.

⁸ It might be argued that the concept of a causal sequence is innate, but this response is not available to the empiricist Ābhidharmikas.

in causal interaction.⁹ How could a series of discrete mental events come to form a concept of a causal sequence if no mental event lasts more than a single moment? The continuity problem, then, is whether *experiential* continuity can be reductively explained in terms of *causal* continuity. If it cannot, the Buddhist reductionist view looks to be self-undermining.¹⁰

The second problem for Buddhist reductionism involves I-consciousness, or first-person experiences and thoughts. If a particular token of the first-person pronoun ‘I’ does not refer to the utterer’s self, what, if anything, does it refer to? It seems obvious that when an individual correctly uses ‘I’, she is referring to herself. But, as the Buddhist reductionist will quickly point out, from this it does not follow that when she uses ‘I’ she refers to a substantial self. Self-reference is not necessarily reference to an ontologically independent self. So even if one rejects the existence of such a self, one can still give an account of first-person self-reference. Perhaps the first-person pronoun is not a genuine referring term, or perhaps, as for Vasubandhu, ‘I’ refers to the continuum in which it occurs, rather than the self. The deeper problem here concerns the centrality and continuity of the sense of self (*ahaṃkāra*) and its connection to self-consciousness, or what Western phenomenologists term ‘the first-person mode of givenness’ of experience.¹¹

The great Nyāya critic of Buddhism, Uddyotakara, presses the difficulty in the following passage:

The consciousness of “I,” which conforms to the distinctions of the nature of the object, and which does not depend upon memory of marks, the possessor of the marks, and their relationship, is direct acquaintance just as is the cognition of physical form. Concerning what you yourself, with perfect confidence, establish to be direct acquaintance, in virtue of what is it that it is [said to be] direct acquaintance? You must establish it as being consciousness alone, which does not depend upon the relationships among marks, etc., and which is self-presenting. So then you think there is an I-cognition, but that its object is not the self? Well, then show us its object! (Kapstein 2002: 98)

As Uddyotakara points out here, first-person self-reference must be anchored in a non-criterial, non-inferential mode of self-acquaintance. But if there is no self, what are I-cognitions directly acquainted with? What is the subject of experience? As Uddyotakara himself mentions, one later Buddhist response to this problem is to argue that consciousness is inherently reflexive or self-presenting (*svasaṃvedana*) and this inherent reflexivity is the basis of both explicit I-cognitions and the more inchoate diachronic sense of self (*ahaṃkāra*). That is, as in Sartre’s view, consciousness is always consciousness of *itself*, but not necessarily consciousness of a *self* (Sartre 1957). Later developments aside, however, it is unclear that the

⁹ There is in fact a debate about whether *dharmas* have only momentary existence, as argued by the Sautrāntikas, or whether, as argued by the Sarvāstivādins, *dharmas* exist in the past, present, and future. In either case, though, *dharmas* are only causally efficacious in the fleeting present. Thus there remains a problem of continuity on either view.

¹⁰ As Ganeri points out, Hume is perhaps more consistent here in denying the (ultimate) reality of causation.

¹¹ ‘First-personal givenness’ is often used interchangeably with ‘subjectivity’. Cf. Zahavi (2005) for a discussion of this.

ruthlessly reductive, impersonalist causalism of the Abhidharma can accommodate the first-personal givenness or the first-personal continuity of human experience.

The third problem for Buddhist reductionism arises from the commitment to mereological reductionism. The properties of a whole, including causal properties, are thought to be reductively determined by the intrinsic properties of its components. Yet the thoroughgoing reductionism of Abhidharma seems to be in tension, with not just non-Buddhist substantialism, but also the dynamic, processual, and multi-level analysis of human beings found in early Buddhism. In early Buddhism, the *skandha*-analysis of the human being should be understood as an *analytical* separation of ontologically interdependent *aspects* of a single body-mind process (*nāma-rūpa-santāna*), rather than as an ontological decomposition of a whole into its independent parts (Hamilton 2000, Garfield 1995: 245). It is primarily an *aspectsual* analysis rather than a mereological one. Furthermore, dependent co-arising (*pratītyasamutpāda*) is a multi-level account of interdependence, not restricted to the micro-level of phenomena. The radical interpretation of dependent origination and anti-substantialism found in Buddhist reductionism may not have the resources required to account for the dependent origination of the human person. And, in any case, as I argue below, there are good reasons to question mereological reductionism, at least with regard to some systems.

The fourth problem for Buddhist reductionism, namely the reification of *dharmas*, is closely related to the third. Recall that Abhidharma ontology rests on a sharp dichotomy between, on the one hand, those entities that have a dependent nature (*parabhāva*) and are therefore merely conventionally real and, on the other hand, those ultimately real entities that have an independent intrinsic nature (*svabhāva*). Clearly mereological reductionism requires an irreducible reduction-base and, as the Ābhidharmikas insist, the entities that form the reduction-base for everyday things must not themselves borrow their nature from other things. That is, macro-level properties, including the properties of wholes, must reductively supervene on the intrinsic, non-relational properties of the base level. Thus ultimately real entities must be independent and basic, as well as individuated by their unique and intrinsic, non-relational properties. Ultimate reality, then, is understood in terms of substance (*dravya*) and essence (*svabhāva*). Therefore, the worry is that this picture constitutes an unwarranted reification of some phenomena (basic *dharmas*) and, at the same time, an unwarranted nihilism about other phenomena (conventional entities). Moreover, as with the Abhidharmika's mereological reductionism, we will argue that there are good reasons to question this reified account of phenomena. Indeed, if the Buddhist Mādhyamikas are right, the Abhidharma view is not just unwarranted, but also incoherent.

The dependent origination of autonomous systems

Given the shortcomings of the Abhidharma theory of persons, what is required is a middle way between their reductionist fictionalism and the substantialism of *ātman* and Cartesian ego theories. Moreover, developments in complex systems theory and biology call into question strongly reductionist approaches such as the Abhidharma. For, unlike flames or chariots (two of the common analogues to the human person),

biological systems display a high degree of self-organized autonomy. On the inactive approach discussed below, living beings are neither enduring substances nor merely aggregative systems, but rather self-regulating unities. Of course, as we have seen, the Abhidharma analysis recognizes that psycho-physical systems are self-perpetuating and characterized by functional integration and feed-back loops. Yet this traditional Buddhist analysis is combined with a strict mereological reductionism that must, in the end, deny genuine causal status to macro-level entities or structures. Therefore, to avoid this problem, in this section I will turn to the theory of autonomous systems, an integral component of the enactive approach, for support in developing an anti-substantialist, but non-reductionist, account of persons.

According to both Buddhist and enactivist accounts, sentient beings are organized dynamic systems. Hence an understanding of the system requires that we pay close attention not just to the system's components, but also to its organization.¹² We may begin with the distinction between *heteronomous* and *autonomous* systems. A heteronomous system is exogenously controlled and can cleanly be modeled as an input-output system. In contrast, an autonomous system primarily will be understood in terms of its "endogenous, self-organizing and self-controlling dynamics," and "does not have inputs and outputs in the usual sense" (Thompson 2007:43). Instead of an input-output model, autonomous systems are understood in terms of perturbation and response. External factors perturb the on-going endogenous dynamics of the system, yielding a response that must be understood in terms of the system's dynamics and its overall organization. More specifically:

In complex systems theory, the term *autonomous* refers to a generic type of organization. The relations that define the autonomous organization hold between processes (such as metabolic reactions in a cell or neuronal firings in a cell assembly) rather than static entities. In an autonomous system, the constituent processes (i) recursively depend on each other for their generation and their realization as a network, (ii) constitute the system as a unity in whatever domain they exist, and (iii) determine a domain of possible interactions with the environment (Thompson 2007: 44; cf. Varela 1979).

In biochemistry, Maturana and Varela (1980) call this type of autonomy 'autopoiesis' (self-production). Autopoiesis involves what Varela terms a 'logical bootstrap' or 'loop' in which a network or process creates a boundary and is subsequently constrained by that boundary. This is the system's *organizational closure* ((ii) above). For instance, at the cellular level, a self-organizing process of biochemical reactions produces a membrane, that, in turn, constrains the process that created it (Varela 2001). The completion of this loop gives rise to a distinct biological entity that maintains its own boundary in its environment. This new level of coherence is a 'virtual identity' that is to be understood in terms of both boundary-maintenance or organizational closure and a new mode of interaction with the environment. In addition, autopoietic systems are characterized by *operational closure* ((i) above): "the property that among the conditions affecting the operation of any constituent process in the system there will always be one or more processes

¹² The following discussion of autonomous systems closely follows Thompson (2007) and Varela (1999, 2001).

that also belong to the system” (Di Paolo 2009: 15). Furthermore, autonomous systems are always coupled to their environments ((iii) above). As Thompson explains, “Two or more systems are coupled when the conduct of each is a function of the conduct of the other” (Thompson 2007: 45). When two systems (organism and environment) develop a history of recurrent interactions leading to a ‘structural congruence’ between them, we have *structural coupling* (Thompson 2007; Maturana 1975; Maturana and Varela 1987).

Sentient beings, on this view, are understood not as heteronomous, mechanical input-output systems, but rather as dynamic, autonomous systems—necessarily coupled to the environment, but also self-controlling. In addition, autonomous systems, in particular living and sentient systems, involve *emergent processes*. As Thompson describes, “An emergent process belongs to an ensemble or network of elements, arises spontaneously or self-organizes from the locally defined and globally constrained or controlled interactions of those elements, and does not belong to a single element” (Thompson 2007: 60). Emergent processes, and the systems in which they arise, exhibit two forms of determination. Local-to-global determination involves the emergence of novel macro-level processes and structures based on changes in the system components and relations. Global-to-local determination involves macro-level processes and structures constraining local interactions. Thus self-organizing systems display *circular causality*: local interactions give rise to global patterns or order, while the global order constrains the local interactions (Haken 1983).

The type of self-production and self-maintenance found in living systems goes beyond the type of self-organization seen in non-living systems. The degree of autonomy found in living beings is, according to the enactive approach, a form of *dynamic co-emergence*.

Dynamic co-emergence best describes the sort of emergence we see in autonomy. In an autonomous system, the whole not only arises from the (organizational closure of the) parts, but the parts also arise from the whole. The whole is constituted by the relations of the parts, and the parts are constituted by the relations they bear to one another in the whole. Hence, the parts do not exist in advance, prior to the whole, as independent entities that retain their identity in the whole. Rather, part and whole co-emerge and mutually specify each other. (Thompson 2007: 65)

A candle flame (a common Buddhist analogy for non-substantial personal continuity) or a Bénard cell, as dissipative systems, will display self-organization and self-maintenance to a degree, but the key boundary conditions that keep these systems away from equilibrium are exogenous. In contrast, in truly autonomous systems, “the constraints that actually guide energy/matter flows from the environment through the constitutive processes of the system are endogenously created and maintained” (Ruiz-Mirazo and Moreno 2004: 238).

Returning to Buddhist reductionism, recall that Vasubandhu’s criteria for the mere conventionality of a phenomenon was its actual or analytical decomposability. Moreover, the issue of decomposability, on the Abhidharma approach, is closely tied to reducibility. Full decomposability requires that the components of a complex

entity are fully specifiable independently of their relations to one another and within the whole. Full reducibility further requires that the properties and (apparent) causal powers of the whole be determined by the intrinsic properties and causal powers of the independent and irreducible components.

However, in complex dynamic systems with *nonlinear* interactions, such as multicellular organisms, the immune system, and the brain, full decomposability is not possible. Nonlinear systems are characterized by non-additive and non-proportional interactions—i.e., nonlinear interactions—and thus the system's properties cannot be aggregatively derived from the properties of its parts (Thompson 2007). As Thompson points out:

An autonomous system is at least minimally decomposable, if not non-decomposable. More precisely, when one adopts an autonomy perspective, one *ipso facto* characterizes the system as at least minimally decomposable. The reason is that an autonomous system is an organizationally and operationally closed network; hence it is the connectivity of its constituent processes that determines its operation as a network. (Thompson 2007: 421).

So, sentient beings, as living autonomous systems, are not amenable to the reductive analysis of the Abhidharma. Sentient beings are not sufficiently decomposable (if decomposable at all) to be exhaustively analyzed and explained in terms of the intrinsic properties and causal powers of independently specifiable components. In addition, the self-organizing, self-maintaining, and self-regulating capacities of living beings rely on both local-to-global and global-to-local influence, and therefore the causal capacities of the system *qua* system are both real and not determined by the intrinsic properties of their most basic components. In the case of autonomous systems such as human beings, we have mereological dependence without strict mereological reduction. On the other hand, it is important to note that the enactive approach is not a return to substantialism. Autonomous systems are not static, ontologically independent substances. Rather the autonomy and irreducibility of living beings derives from dense networks of relationality and interdependence. That is, autonomous systems are dependently originated (*pratītyasamutpāda*).

Emptiness and the virtual self

A turn toward the enactive approach and its autonomy perspective can help to find a middle way between substantialism and reductionism about persons. Persons can be understood in dynamic-relational terms as autonomous systems. So far, however, the focus has been on persons as sentient beings—that is, as embodied and embedded biological systems. I will now turn to the importance of the deeply entrenched human sense of self. In addition, just as I have used the enactive approach to expand upon and modify a Buddhist analysis, I will in turn use later developments in Buddhist thought (in particular the Madhyamaka school) to correct what I take to be shortcomings in Francisco Varela's enactivist account of the self.

Varela (1999, 2001), explicitly drawing on Buddhist philosophy, argues that the human self is both emergent and *virtual* or *empty* (*śūnya*). He therefore rejects the existence of a substantial, bounded, enduring self. The self, he argues, emerges from

the human organism's endogenous neurobiological dynamics and from its embeddedness in its natural and social-linguistic environment. Thus we create and re-create ourselves from moment to moment through the dynamic interaction of brain, body, language, and world. He writes:

Why do emergent selves, virtual identities, pop up all over the place creating worlds, whether at the mind/body level, the cellular level, or the transorganism level? This phenomenon is something so productive that it doesn't cease creating entirely new realms: life, mind, and societies. Yet these emergent selves are based on processes so shifty, so ungrounded, that we have an apparent paradox between the solidity of what appears to show up and its groundlessness (Varela 2001).

These systems behave *as if* a central agent or controller is directing them—yet no such central agent can be found. This is what Varela means by the virtual or empty self: “a coherent global pattern that emerges from the activity of simple local components, which seems to be centrally located, but is nowhere to be found, and yet is essential as a level of interaction for the behavior of the whole” (Varela 1999: 53).

Of course, it is natural to think that while an ant colony or an amoeba may have only a *virtual* self, surely we humans are the real deal. But Varela is awowedly in agreement with the Buddhist theory of non-self. On his account, “either we are unique in the living and natural world, or else our very immediate sense of a central, personal self is the same kind of illusion of a center, accountable by more of the same kind of analysis [i.e., in terms of autopoiesis]” (Varela 1999: 61). Moreover, “what we call ‘I’ can be analyzed as arising out of our recursive linguistic abilities and their unique capacity for self-description and narration” (Varela 1999: 61). Indeed, this linguistically constructed self serves as what he terms a ‘virtual interface’ between the body and the natural and social environment in which it is embedded.

Now the virtuality or emptiness of the self can be taken in a weaker or a stronger sense, and it is not always clear which sense Varela intends. In the weaker sense, the emergent self is virtual merely because it is distributed or not localized, and thus insubstantial. The self is not an illusion, but its supposed singularity and localizability is a fiction or projection. This would rule out naive homuncular or substantialist accounts of the self without entailing fictionalism about the self *per se*. In the stronger sense, the self is virtual in that it is an illusion or useful fiction. The insect colony behaves *as if* someone were in charge, *as if* it had a unified perspective, and so on, but this is an illusion. Likewise, our sense of self, our sense of having (or being) a unified first-person perspective, and being a center of agency are also illusions.

Despite some ambiguity in Varela's characterization of the virtual self, I take him to be arguing for the stronger sense of virtuality. First, he repeatedly points out the ways in which mainstream findings in cognitive science challenge the notion of the self as a unified and coherent point of view—i.e., not just inner homunculus *explanations* of a unified perspective, but also the very existence of such a perspective is challenged (Varela 1999: 36–41). Second, Varela notes that he and Daniel Dennett—the arch-fictionalist—have come to the same conclusions regarding the self (Varela 1999, 2001). Thus, on Varela's view, the self is a virtual or fictional

construct that emerges from the distributed activity of a natural, autopoietic system and, in the case of the human self, the system's use of language and its embeddedness in a linguistic community.

Now the resonance between Varela's account and Buddhism should be obvious. And, of course, Varela draws heavily from Buddhist ideas and practices in the formulation and defense of his account of the virtuality of the self, especially the concept of emptiness (*śūnyatā*) as developed in the Madhyamaka, or Middle Way, school of Buddhism. Indeed, he argues that the emptiness of the self "is the golden thread that unites our self-understanding with an external and scientific account of mental functioning" and further that ethical wisdom rests on firsthand acquaintance with the empty nature of the self (Varela 1999: 36). However, a proper understanding of emptiness casts doubt on Varela's account of the self as merely virtual.

As mentioned at the outset of this paper, the account of human beings as selfless persons (*pudgalanairātmya*) is held by all major Buddhist schools, but there has been a great deal of disagreement as to the full ontological implications of the rejection of a substantial self. In contrast to the reductionists, Mādhyamika thinkers allow for an ontologically deflationary account of the self. The self is said to have an experiential and practical reality, while still insisting that this minimal self is not to be reified. As the Dalai Lama explains the latter view, "both body and mind are things that belong to the I, and the I is the owner; but, aside from mind and body, there is no separate independent entity of I. There is every indication that the I exists; yet, under investigation, it cannot be found" (Gyatso 2000: 65). This minimal self—what is called the 'mere I' or 'mere self' (Tibetan: *nga tsam*)—is an emergent phenomenon that, while real, is not a substantial separate *thing*, and therefore disappears under analysis.¹³

The difference between the Abhidharma reductionist fictionalism and the later Buddhist deflationary non-reductionism turns on competing accounts of the concept of emptiness. To be empty, on the reductionist view, is to lack ontological independence and an intrinsic nature and, therefore, to be nothing more than a conceptual construct or convenient fiction. In contrast, the Madhyamaka school—which Varela, Thompson, and Rosch claim as one basis of their thought in *The Embodied Mind* (1991: 217–235)—takes a different approach to the concept of emptiness. The Mādhyamikas agree that to be empty is to lack *svabhāva* or inherent existence, but go on to argue that the notion of *svabhāva* is itself untenable. Rather than arguing that conventional phenomena can only be accounted for in terms of ultimate phenomena, the Mādhyamika argue that positing ultimate, non-empty phenomena actually *precludes* a coherent account of the conventional world. The Mādhyamikas, therefore, argue for the selflessness of persons (*pudgalanairātmya*) and the selflessness (i.e., emptiness) of all phenomena (*dharmanairātmya*).

But what does it mean to say that all things are empty, that they lack *svabhāva*? Of course, within the Abhidharma framework, to say this amounts to nihilism. To be conventionally real is to be purely derivative, indeed to be a mere convenient fiction.

¹³ The analysis here is the type of ontological analysis that looks for the substantial reality of the object.

Yet, how could *everything* be a mere convenient fiction? Within the Mādhyamika framework, in contrast, for a thing to be empty is not for it to be unreal, but rather for its existence and nature necessarily to depend on other things. There is a three-way implication between emptiness, dependent origination, and conventional reality. As Jay Garfield explains:

When we say that a phenomenon is empty, we mean that when we try to specify its essence, we come up with nothing. When we look for the substance that underlies the properties, or the bearer of the parts, we find none. When we ask what it is that gives a thing its identity, we stumble not upon ontological facts but upon conventions. For a thing to be non-empty would be for it to have an essence discoverable upon analysis; for it to be a substance independent of its attributes, or a bearer of parts; for its identity to be self-determined by its essence. A non-empty entity can be fully characterized nonrelationally. (Garfield 2002: 38)

So, according to the Mādhyamikas, anything that exists depends on other things for its existence and nature, and depends (in part) on our practices of individuation for its identity-conditions.¹⁴

Returning to the question of the self, one can see why Varela's identification of virtuality with emptiness is problematic. On the Mādhyamika account of emptiness, which Varela himself endorses, emptiness simply does not entail virtuality in either its weaker or stronger senses. To call the self virtual is to imply that it is either unreal or less real than other things. But showing that the self emerges from and depends on lower-level processes, that it is not an independent substance, that in trying to specify its identity-conditions we make reference to our interests and practices, or that it has no absolute ontological primacy does not cast doubt on its existence. Rather, it shows that the self is empty, *just like everything else*. To think otherwise, on the Mādhyamika view, is to accept an ontological foundationalism and essentialism that they argue is incoherent.¹⁵

The larger point here is that, insofar as all phenomena are constitutively embedded in a network of relations (causal, mereological, emergence, etc.), there is nothing especially *virtual* about the emergent self. In aligning his account of the emptiness of the self with the fictionalist views of the Abhidharma school (Varela, et al 1991: 58–81) and Dennett, Varela implicitly reifies lower levels of stable organization while simultaneously negating the conventional reality of the emergent self. In contrast, Thompson is in agreement with the Mādhyamika school when he insists that, “Phenomena at all scales are not [independent] entities or substances but relatively stable processes, and since processes achieve stability at different levels of complexity, while still interacting with processes at other levels, all are equally real and none has absolute ontological primacy” (Thompson 2007: 441). To claim, as Varela seems to do, that the emergent global pattern is virtual, but that the

¹⁴ Hence, the Mādhyamika holds that all phenomena are constitutively interdependent and also rejects the idea of the ‘ready-made world’ characteristic of metaphysical realism.

¹⁵ See Garfield (1995), Siderits (2003), and Westerhoff (2009) for in-depth discussions of Madhyamaka arguments.

components from which it emerges are actual, is to miss the full implications of the dynamic-relational ontology at the heart of both Buddhism and the enactive approach.

The minimal self

Mādhyamikas argue that the self is empty, but that it is neither a mere fiction, nor reducible to the body-mind continuum.¹⁶ The great Tibetan Madhyamaka philosopher, Tsongkhapa, distinguishes three approaches to the question of the self: (1) the substantialism of non-Buddhist schools, (2) Buddhist reductionism, and (3) the view of his tradition of Prāsangika Madhyamaka (Jinpa 2002: 109). The first type of view holds that the self is a kind of entity that exists independently of the *skandhas* and combines non-reductionism and realism. The second type of view holds that the self does not exist at all and that persons are fully reducible to the *skandhas*. Terms such as ‘I’, ‘self’, and proper names in fact refer to the impermanent mental and physical elements. This type of view, according to Tsongkhapa, combines realism (about the elements as the reduction base) with reductionism. Finally, the Prāsangika Madhyamaka school holds that the self is dependent upon, but not reducible to the *skandhas*. This view combines ontological deflationism (i.e., all things are empty of inherent existence) with non-reductionism. As Tsongkhapa writes:

There are two senses of the term ‘self’: a self conceived in terms of intrinsic nature that exists by means of intrinsic being, and a self in the sense of the object of our simple, natural thought ‘I am.’ Of these two, the first is the object of negation by reasoning, while the second is not negated, for it is accepted as conventionally real. (Jinpa 2002: 71)

Unlike their reductionist forbearers, therefore, the Mādhyamikas accept the *minimal self* as the object of the natural sense of self or I-consciousness. Here we see the Madhyamaka answer to Uddyotakara’s challenge to identify the referent of I-consciousness. I-thoughts *do* refer to the self rather than a mere bundle of physical and mental events, but this minimal self is not an enduring substance.

But once one has rejected the substantialist account of the self, why not identify the self with the psycho-physical continuum? According to the Mādhyamikas, the self cannot be identified with the aggregates because, firstly, they have incompatible properties. The self is, by hypothesis, single and persisting, while the aggregates are multiple and fleeting.¹⁷ Identifying them would lead to a multiplication of selves within the life of single person—a result taken to be absurd. Moreover, the self and the aggregates have different persistence-conditions, and therefore cannot be identical. Secondly, it is argued that memory presupposes the continuity of a first-person perspective or I-consciousness that cannot be accounted for in impersonal, reductionist terms. That is, the first-person perspective needed to account for genuine

¹⁶ The Mādhyamikas do say that the self is illusion-like in that it appears to have substantial existence, but is in fact empty. But, again, according to this view *all* phenomena are illusion-like in this sense.

¹⁷ Mādhyamikas reject the ‘pearl’ view of the self (cf. Strawson 1999)—the view that the person consists of series of short-lived selves-and maintain, on phenomenological and pragmatic grounds, that the self is persistent. But, since the self is a process, it perdures rather than endures.

memory disappears from the impersonal causalism of reductionism.¹⁸ Thirdly, Tsongkhapa argues that our conventions and practices, such as assigning moral responsibility, take the notion of a person and the ‘mere I’ as basic (Jinpa 2002). Finally, the Mādhyamikas argue that both substantialist and reductionist approaches to the self fail to account for the inherently indexical, perspectival nature of the minimal self. Aryadeva argues: “That which is self to you is not self to me; from this fixed rule it follows that that is not self. Indeed, the construction (*kalpanā*) [of a sense of self] arises out of the impermanent things” (Ganeri 2007: 191). Candrakīrti expands on this argument in his commentary on Aryadeva:

That which is self to you, the focal point of your sense of ‘I’ (*ahaṃkāra*) and self-interest (*ātmasneha*), that indeed is not self to me; for it is not the focal point of my sense of ‘I’ and self-interest. 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 invariably present. One should give up the superimposition of [such] a self, for it is something the content of which is unreal (*asadartha*). (Ganeri 2007: 192)

Having a sense (or concept) of self entails being able to draw a distinction between self and other, and to experience things as mine and not mine. However, substantialist or entitative views of the self, according to Aryadeva, cannot ground this indexical, perspectival distinction because they take the self to be a kind of *thing*.

Yet the reductionist view, despite denying the existence of the self, fares no better. The project of Buddhist reductionism is to account for persons in impersonal terms, that is, in terms of causal connections between fleeting mental and physical events. In shifting to an impersonal and non-perspectival standpoint, the reductionist loses sight of the first-person perspective and it is unclear in this case how to derive the perspectival from the non-perspectival. There appears to be an explanatory gap between the first- and third-person standpoints. Of course, for the reductionist, the fact that the minimal self is not found when one shifts to a non-perspectival, third-person discourse is grounds to deny its real existence. However, the Mādhyamikas deny that this discourse has any absolute metaphysical or explanatory priority. The problem with reductionist accounts of personal identity, then, is that they deny the importance of the first-person perspective.

One helpful and important aspect of the Mādhyamika approach is that it shifts the discussion from a concern with third-person, metaphysical issues about the existence of selves and persons to a concern with the first-person, lived sense of self (*ahaṃkāra*). The attempt to find a metaphysical *ground* for the existence of the self either in a mental substance or in a reduction to impersonal aggregates is eschewed by the deflationist Mādhyamikas as merely two instances of the error of reification. Instead, the Mādhyamika account of the empty minimal self is given in terms of its experiential and practical reality. It is fundamentally a matter of the structure and continuity of the first-person perspective.

¹⁸ On this point, the Mādhyamikas agree with the Nyāya critics of Buddhist reductionism.

In discussing the Mādhyamika approach to the minimal self Thupten Jinpa remarks:

One of the fundamental premises of Tsongkhapa's thought . . . is that an individual's sense of self, or I-consciousness, is innate. It is instinctual and natural. It is neither linguistic nor even conceptual, if by conceptual one presupposes [reflective] self-awareness. It is a natural, reflexive consciousness, almost like an underlying sense of one's own existence. (Jinpa 2002: 123)

Our most basic sense of self is pre-conceptual, pre-linguistic, and natural.¹⁹ But we must proceed with caution here. Despite holding that the sense of self is natural and pre-linguistic, Tsongkhapa denies that each moment of consciousness is self-presenting (*svapṛakāśa*). Like Mādhyamikas such as Candrakīrti and Śāntideva, he refutes the idea that all consciousness inherently involves pre-reflective self-consciousness (*svasaṃvedana*).²⁰ The 'mere I' (Tibetan: *nga tsam*) here refers to the *ahaṃkāra*, not *svasaṃvedana*.²¹ Here again, Tsongkhapa's view of the self differs from Varela's. Recall that, on Varela's account, the 'I' arises from our "recursive linguistic abilities and their unique capacity for self-description and narration" (Varela 1999: 61), whereas the minimal self in Tsongkhapa's account is experientially prior to linguistic construction.²²

Furthermore, on Tsongkhapa's view, the minimal self has a diachronic dimension.²³ He explains:

The self that is the focus of Devadatta's instinctual sense 'I am' when not thinking of a specific temporal stage [of his existence] is the mere I that is within him since beginningless time. The individual selves [of Devadatta] when he appropriated the body of a celestial being and so on are only instances of the former [mere I]. Therefore, when an I-consciousness arises in Devadatta focusing specifically on a particular form of existence [e.g., as a human], the object of his I-consciousness is a particular instance of Devadatta's self. (Jinpa 2002: 123)

The idea here is that the mere I is not confined to the present, but rather provides a basic form of continuity throughout the different phases or temporal stages of one's life. Indeed it is precisely the continuity of this basic first-person perspective that explains why the various forms of existence are parts of the same life history. Each of the 'individual selves' is based upon the minimal self; each tokening of an I-thought is a particular instance or expression of this mere I.

Tsongkhapa further argues that the minimal self allows us to explain the coherence of our personal plans and projects. When we plan for the future or undertake a particular project, he argues, we do not "make distinctions between the

¹⁹ Thompson defines sentience as "the feeling of being alive" (2007: 161), while neuroscientists Damasio (1999) and Panksepp (1998) posit a primitive 'feeling of self'.

²⁰ See Dreyfus (1996), Garfield (2006), MacKenzie (2007), MacKenzie (2008), and Williams (1997) for further discussion.

²¹ Unlike, Tsongkhapa, I *do* accept the notion of *svasaṃvedana*. On my view, the minimal self (*ahaṃkāra*) emerges from the more basic inherent reflexivity of consciousness. Thus my view is closer to the Mādhyamika of Śāntarākṣita (in India) or the Kagyu and Nyingma traditions (in Tibet).

²² This is not to say that linguistic construction plays no role in Tsongkhapa's account, but only to point out that the minimal self is pre-linguistic.

²³ The Madhyamaka account of the minimal self, therefore, differs from Antonio Damasio's notion of the core self in that the core self has no long-term temporal extension.

self of this time or that time. Rather, these endeavors are motivated by the simple wish for the self to be happy and overcome suffering. And since the self as a generality does pervade all temporal stages [of a person's existence], these acts also cannot be said to be deluded" (Jinpa 2002: 124). For the Prāsangika Mādhyamika such as Tsongkhapa, these pragmatic, experiential considerations concerning the self are central.

Self-appropriation

On the Mādhyamika account, the empty self is neither independent of nor reducible to the five *skandhas*. What, then, is the dependence relation between the pre-personal *skandha-santāna* and the self? A common analogy for the relation between the self and the *skandhas* is fire and fuel. Just as the fire appropriates (*upādāna*) the fuel to perpetuate itself, the self appropriates as its own the various mental and physical events that make up the *skandha-santāna*. As Candrakīrti comments on Nāgārjuna's use of the analogy:

That which is appropriated is the fuel, the five [types of] appropriated element. That which is constructed in the appropriating of them is said to be the appropriator, the thinker, the performing (*niṣpādaka*) self. In this is generated [the activity of] 'I'-ing, because from the beginning it has in its scope a sense of self. (Ganeri 2007)

The self, then, is the appropriator (*upādātṛ*) and the various elements are the appropriated (*upādāna-skandha*), and yet Candrakīrti insists 'the self is not a real, existent thing'. That is, the self lacks inherent existence (i.e., it is empty) and it is not any kind of thing or object. Rather, the self is 'I'-ing (*ahaṃmāna*) or on-going self-appropriative activity (Ganeri 2007). Furthermore, 'I'-ing is an inherently perspectival activity; it appropriates phenomena as 'me' and 'mine', incorporates them into its own on-going dynamic, by *indexing* (or tagging) them to the I. Appropriation, then, functions as a self-referential loop.

According to Jonardon Ganeri, the Mādhyamika theory is a *performativist* theory of the self (2007). On his interpretation,

When I say 'I am in pain,' I do not *assert* ownership of a particular painful experience; rather, I *lay claim* to the experience within a stream. This is a performativist account of the language of the self, in which 'I' statements are performative utterances, and not assertions, and the function of the term 'I' is *not* to refer. (Ganeri 2007: 202)

Of course, Ganeri's performativist reading is congenial to my enactivist account of the self, but it differs in two ways. First, Ganeri's interpretation relies on a distinction between referential and performative utterances that is sharper than I would want to endorse.²⁴ Second, while I agree with Ganeri (and Varela) that

²⁴ One can reject an entitative view of the self and still allow that self-talk is referential. In this case, self-talk refers to self-appropriating activity (and thereby also to that which is appropriated). Further, the use of a term can be both performative and referential, as when one says "I now pronounce you husband and wife." Here 'pronounce' is both performative and referential.

language is central to the full constitution of the self, I also agree with Tsongkhapa that the minimal self has its roots in the pre-linguistic structure of lived experience.

More specifically, the root of the minimal self is the *recursive* nature of lived experience. A recursive process is one wherein the results of the process are fed back into the process itself. On the Buddhist view, the vicious cycle of *saṃsāra* is understood in terms of the recursive process of dependent origination. Indeed, living itself is a recursive process. As Hans Jonas remarks, “organisms are entities whose being is their own doing ... the being that they earn from this doing is not a possession they then own in separation from the activity by which it was generated, but is the continuation of that very activity itself” (Jonas 1996: 86). In order to survive, the organism must maintain its own dynamic organization in the face of, but also in virtue of, continuous matter-energy turnover. The viable organism, through its organizational and operational closure, is able to subsume or appropriate both bits of the environment and elements of the organism itself. Thus, ‘I’-ing is perhaps more like the process of metabolizing than it is like combustion.

In addition to the recursivity of biological processes, the stream of experience has its own recursive structure.²⁵ The ‘stream’ of experience is a temporal flow. Yet, as Śrīdhara pointed out in his criticism of the Abhidharma, it is hard to see how a series of discrete momentary experiences could constitute the type of on-going point of view required to even form the concept of a causal sequence. Furthermore, the problem involves not only external objects and processes, but also how we can be aware of our own experiences as forming a unified temporal flow. Without this, we will be unable to account for the emergence of the minimal self through self-appropriation. What is required is an account of how impressions are retained within the temporal flow of experience—that is, we must have an account of time-consciousness (Husserl 1991).²⁶

The basic unit of temporal experience for Husserl (as for James) is not a durationless point, but rather a moment with temporal thickness. The structure of this ‘duration-block’ is *protention-primal impression-retention*. As Husserl explains:

In this way, it becomes evident that concrete perception as original consciousness (original givenness) of a temporally extended object is structured internally as itself a streaming system of momentary perceptions (so-called primal impressions). But each such momentary perception is the nuclear phase of a continuity, a continuity of momentary gradated retentions on the one side, and horizon of what is coming on the other side: a horizon of “protention,” which is disclosed to be characterized as a constantly gradated coming. (Husserl 1977: 154)

The primal impression is restricted to the now-phase in a sequence. In listening to a melody, the primal impression is directed to the currently sounding note. Retention is directed toward the just-elapsed note. The elapsed note is not actually present in consciousness, but is retained intentionally. Protention is directed toward the future,

²⁵ Though of course, we do not want to make too sharp a distinction between the biological and the experiential here. These are two aspects of one process of *living*.

²⁶ See Zahavi (2005) for an illuminating discussion of these issues.

the next note about to be heard. Whereas the currently sounding note is given in the vivid immediacy of the present, and the just past note is determinately retained, the upcoming note is not given in a fully determinate manner.

This three-fold structure forms a unified whole, the continuous operation of which allows for the experience of temporal continuity. The structure constitutes the living present within which temporal experience ‘wells-up’. Further, on Husserl’s view, the primal impression-protention-retention structure of consciousness accounts for the temporal unification of the stream of consciousness itself. Retention retains the prior phases of the stream, while protention reaches out toward future moments of consciousness. It is through this process, which Husserl calls *horizontal intentionality*, that consciousness is self-affecting or temporally given to itself. Furthermore, horizontal intentionality makes possible what Husserl calls *transverse intentionality*. It is the transverse intentionality of time-consciousness that allows for the continuous experience of a temporal object, such as melody or a spoken sentence. Because the now-phase of consciousness takes an object (e.g., a note) and is retained in the stream, so too is the object of the now-phase of consciousness. In sum, the three-fold structure of time-consciousness is the condition of the possibility of both the diachronic unification of the stream of consciousness and the experienced continuity of temporal objects.

Husserl’s analysis of time-consciousness shows that consciousness is itself recursive. Consciousness takes in its impressions and retains them, marking the impression as past and making the past impression available for the on-going flow of consciousness. Indeed the process of retention is iterative in that not only ‘pastness’, but the degree of ‘pastness’ is marked within the flow of experience. The temporal flow of consciousness involves retentions of retentions, thereby allowing the experience of a temporal *sequence*. Moreover, this recursive process is self-referential. As James Mensch observes:

In retention the subject does not just have the experience of the retained, it experiences itself having this experience, i.e., as retaining the retained. Accordingly, when it grasps an object through a series of retained contents, it prereflectively grasps itself in its action of retention. This grasp is a grasp of itself as having experience, i.e., of itself as a subject. Such self-experience implies that the self-referential character of retention grounds the subject as nonpublic, i.e., as referring (or being present) only to itself. (Mensch 2001: 107)

The upshot of these brief Husserlian considerations is that self-making (*ahaṃkāra*) (what Tsongkhapa calls the object of our “simple, natural thought ‘I am’”) is grounded in the recursive temporality of the stream of consciousness (*vijñāna-santāna*). Moreover, this analysis of experiential continuity answers Śrīdhara’s objection. Even without a substantial self, one can form the concept of a causal sequence because the stream of experience is characterized by horizontal and transverse intentionality. The root of the minimal self, then, is not in linguistic self-appropriation, but in temporal self-appropriation. The ‘I’, or rather ‘I’-ing, emerges from self-grasping and in this sense we are, in Ganeri’s phrase, “whirlpools of self-appropriating action” (2007: 204).

Yet, while temporal self-appropriation is necessary for a diachronically extended minimal self, it is not clear that it is sufficient. What else might be required? One

type of view holds that language or concepts are required for the emergence of even the minimal self. Another type of view, in contrast, holds that pre-linguistic awareness of the body and action are required. In support of this latter view, one may cite, for instance, recent research on neonatal imitation. As Shaun Gallagher summarizes the significance of this research:

Neonates less than an hour old are capable of imitating the facial gestures of others in a way that rules out reflex or release mechanisms, and that involves a capacity to learn to match presented gesture. For this to be possible the infant must be able to do three things: (1) distinguish between self and non-self; (2) locate and use certain parts of its own body proprioceptively, without vision; and (3) recognize that the face it sees is of the same kind as its own face (the infant will not imitate non-human objects). One possible interpretation of this finding is that . . . the human infant is already equipped with a minimal self that is embodied, enactive and ecologically tuned. (Gallagher 2000: 18)

Furthermore, as ecological psychologists point out, the visual field affords information about the perceiver's environment as well as self-specifying information. In seeing the cup, one is also gaining information about one's own position in relation to the cup, information that is crucial to coordinating perception and action. Thus, on this account, the minimal self is embodied and enactive, as well as temporal.

Conclusion

The Buddhist-enactivist conception of the self explored here provides, I argue, a middle path between substantialism and reductionism, between treating the self as either an independent entity or a mere fiction. The fundamental problem for substantialism is *change*: positing a fundamentally unchanging substratum of our identity in the face of incessant change ultimately alienates the self from experience and the world. Nothing in our experience—neither body, nor mind, nor world—has the kind of permanence and stability attributed to the substantial self. One might take the substantial self as an explanatory posit, but it remains unclear whether this posit is necessary or even whether it could do any real work.

On the other hand, the fundamental problem for reductionism is the centrality of the first-person perspective. A living sentient being is not a mere aggregate or bundle of components and cannot fully be understood from a purely external, third-person perspective. The living organism displays the interiority of its own immanent purposiveness and its needful and precarious relations to its (enacted) milieu. The stream of experience becomes self-referential, given to itself, through the recursive structure of time-consciousness. The embodied being is pre-reflectively aware of itself in and through its active, striving body. (Thompson 2007) The human person makes her life intelligible and accountable to herself and others through the self-referential resources of language and narrative (Zahavi 2005). In each of these ways (and others) the phenomena of subjectivity resists reduction. The common themes here are recursivity, self-organization, and self-reference. The key to understanding (or beginning to understand) self-making, I suggest, is to see how dynamic processes enact themselves through self-organizing, self-appropriating activity—biologically,

experientially, and socially. For, if the Buddhists and the enactivists are right, we are this activity.

References

- Albahari, M. (2006). *Analytical Buddhism: The two-tiered illusion of self*. Houndsmills, NY: Pallgrave Macmillan.
- Arnold, D. (2005). *Buddhists, Brahmins, and Belief: Epistemology in South Asian philosophy of religion*. New York: Columbia University Press.
- Damasio, A. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. New York: Harcourt Brace.
- Dennett, D. (1991). *Consciousness explained*. Boston: Little, Brown.
- Di Paolo, E. (2009). Extended life. *Topoi*, 28(1), 9–21.
- Dreyfus, G. (1996). *Recognizing reality: Dharmakīrti's philosophy and its Tibetan interpretations*. Albany: State University of New York Press.
- Duerlinger, J. (2003). *Indian Buddhist theories of persons: Vasubandhu's "Refutation of the Theory of a Self"*. London: Routledge/Curzon.
- Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Science*, 4(1), 14–21.
- Ganeri, J. (2007). *The concealed art of the soul: Theories of self and practices of truth in Indian ethics and epistemology*. Oxford: Oxford University Press.
- Garfield, J. (1995). *Fundamental wisdom of the middle way: Nāgārjuna's Mūlamadhyamakārikā*. Oxford: Oxford University Press.
- Garfield, J. (2002). *Empty words: Buddhist philosophy and cross-cultural interpretation*. Oxford: Oxford University Press.
- Garfield, J. (2006). The conventional status of reflexive awareness: what's at stake in the Tibetan debate? *Philosophy East and West*, 56(3), 201–228.
- Gyatso, T. (2000). *The meaning of life: Buddhist perspectives on cause and effect*. Somerville: Wisdom.
- Haken, H. (1983). *Synergetics: An introduction*. Berlin: Springer-Verlag.
- Hamilton, S. (2000). *Early Buddhism: A new approach: The I of the beholder*. London: Routledge/Curzon.
- Holder, J. (2006). *Early Buddhist discourses*. Indianapolis: Hackett Publishing Company.
- Husserl, E. (1977). *Phenomenological psychology: Lectures, summer semester, 1925, trans. John Scanlon*. The Hague: Martinus Nijhoff.
- Husserl, E. (1991). *On the Phenomenology of the Consciousness of Internal Time (1893–1917), trans. J. B. Brough*. Dordrecht: Kluwer Academic.
- Jinpa, T. (2002). *Self, reality, and reason in Tibetan philosophy: Tsongkhapa's quest for the middle way*. London: Routledge/Curzon.
- Jonas, H. (1996). *Mortality and morality: A search for good after Auschwitz*. Evanston, IL: Northwestern University Press.
- Kapstein, M. (2002). *Reason's traces: Identity and Interpretation in Indian and Tibetan Buddhist thought*. Boston: Wisdom.
- Lusthaus, D. (2002). *Buddhist phenomenology: A philosophical investigation of Yogācāra Buddhism and the Ch'eng Wei-shu lun*. London: Routledge/Curzon.
- MacKenzie, M. (2007). The illumination of consciousness: approaches to self-awareness in the Indian and western traditions. *Philosophy East and West*, 57(1), 40–62.
- MacKenzie, M. (2008). Self-awareness without a self: Buddhism and the reflexivity of awareness. *Asian Philosophy*, 18(3), 245–266.
- Maturana, H. (1975). The organization of the living: a theory of the living organization. *International Journal of Man-Machine Studies*, 7(3), 313–332.
- Maturana, H., and Varela, F. (1980). *Autopoiesis and Cognition: the Realization of the Living*. Boston Studies in the Philosophy of Science, vol. 42. Dordrecht: D. Reidel.
- Maturana, H., & Varela, F. (1987). *The tree of knowledge: The biological roots of human understanding*. Boston: Shambhala Press/New Science Library.
- Mensch, J. (2001). *Postfoundational phenomenology: Husserlian reflections on presence and embodiment*. University Park, PA: Pennsylvania State University Press.

- Metzinger, T. (2003). *Being no one: The self-model theory of subjectivity*. Cambridge, MA: MIT.
- Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. Oxford: Oxford University Press.
- Ruiz-Mirazo, K., & Moreno, A. (2004). Basic autonomy as a fundamental step in the synthesis of life. *Artificial Life, 10*, 235–259.
- Sartre, J.-P. (1957). *Transcendence of the ego: An existentialist theory of consciousness*, trans. F. Williams and R. Kirkpatrick. New York: Noonday.
- Siderits, M. (2003). *Personal identity and Buddhist philosophy: empty persons*. London: Ashgate.
- Strawson, G. (1999). The Self. In S. Gallagher & J. Shear (Eds.), *Models of the self* (pp. 1–24). Thorverton: Imprint Academic.
- Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Cambridge, MA: Harvard University Press.
- Varela, F. (1979). *Principles of biological autonomy*. New York: ElsevierNorth Holland.
- Varela, F. (1999). *Ethical know-how: Action, wisdom, and cognition*. Stanford: Stanford University Press.
- Varela, F. (2001) The Emergent Self. In *Edge*, vol. 86, www.Edge.org. Accessed 27 February 2006.
- Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT.
- Zahavi, D. (2005). *Subjectivity and selfhood: Investigating the first-person perspective*. Cambridge, MA: MIT.
- Westerhoff, J. (2009). *Nāgārjuna's Madhyamaka: A philosophical introduction*. Oxford: Oxford University Press.
- Williams, P. (1997). *The reflexive nature of awareness: A Tibetan Madhyamaka defence*. London: Routledge/Curzon.