

3

Nirvana and Ownerless Consciousness

MIRI ALBAHARI

1. Introduction

A Buddhist friend remarked recently: ‘Perhaps one or two *arahants* exist in the world’. ‘Arahant’ (or *arhat*) is a Buddhist term for someone who has attained the *summum bonum* of Buddhist practice. Such a state is known as ‘enlightenment’ or *nirvāna*. While Buddhist traditions will differ in their exact depictions of nirvana, most would agree that a sense of the self, with its attendant feelings of ‘me’ and ‘mine’, is extinguished. With it is extinguished the capacity to suffer mentally. There is a radical shift in motivational structure: no longer do such persons seek gratification from any state of affairs. Losing family or suffering illness fails to dent their equanimity. The arahant operates from a different basis: no more identifying with the ‘I’ of such situations than most of us would identify with burning leaves on a fire. Yet they still act fluently in the world—with great joy and spontaneity and compassion.

The ‘perhaps’ of my Buddhist friend was meant to indicate the extreme rarity of such people. Buddhist tradition holds the pull of craving and attachment, needed to sustain the illusion of self, to be so strong that it would take lifetimes of dedicated practice to vanquish. As a philosopher, I am interested in taking the ‘perhaps’ another way: not as an indication of rarity, but of *modality*. Is it really psychologically possible for an arahant to exist, human brains and minds being what they are? Can people really become so free from the sense of self that they no longer identify with their bodies or minds, and yet still act fluently and without suffering in the world? Or is such

an idea, which has inspired thousands to be ordained as monks or nuns, likely to be steeped in religious fantasy? Serious investigation into the possibility of nirvana has not yet entered mainstream analytic philosophy, even though its implications for the metaphysics of mind, if it *were* possible, could well be significant. Western philosophy has tended to zero in on the structure of the ordinary person's mind, the extraordinary being confined to pathological impairment. If nirvana was shown to be possible, then I contend that much of value could be learnt from analyzing the mind's extraordinary capacities.

To the extent that nirvana can be described (itself a matter of contention), it is so multifaceted that no investigation could do justice to all aspects within a single work. This is compounded by the fact that Buddhist traditions diverge on how nirvana is to be exactly understood. My own interpretation focuses particularly on the relation between nirvana and no-self, and is based upon a philosophical reconstruction from discourses (sutras) of the Pali Canon (which I call 'Pali Buddhism' for short). We will see that this challenges an orthodoxy considered by many to be classically 'Buddhist'.

I mentioned that most Buddhist traditions agree that nirvana entails insight into the truth of no-self, such that certain feelings of 'me' and 'mine' are vanquished (however these key notions are to be understood). So when investigating the possibility of nirvana, it is worth asking of a chosen Buddhist tradition: What is the most likely relationship between the sense of self and the ubiquitous feeling of ownership (or 'mineness') had towards one's thoughts and experiences? Could any form of consciousness survive the possible destruction of these ownership-feelings? Could ownerless consciousness be an underlying feature of the everyday mind? The goal of this paper is to draw together the most salient points about consciousness, ownership, and no-self that arose from my initial foray into Pali Buddhism, and to further develop some of the arguments in relation to later work on the subject.¹ I hope it will transpire that even a preliminary investigation into the psychological possibility of nirvana can be a valuable exercise in drawing distinctions that may help to illuminate the architecture of the *everyday* mind.

The paper has three parts. In Part One, I outline in some detail the notion of self that I contend is most central to Pali Buddhism: the self that is

¹ Many of the concepts and arguments in this paper originated in my book *Analytical Buddhism* (2006) and in my (2009) paper, some of which are developed here further, sometimes from a different angle. Other arguments (such as that supporting my interpretation of no-self in Pali Buddhism) are entirely new. (A detailed canonical defense of my interpretation of nirvana and no-self can be found in chapters 2–3 of Albahari 2006).

purported to have illusory status. It is not an abstract or ethereal concept (such as a non-physical soul), but a notion that describes the very thing that most of us unquestioningly take ourselves to be. If this were indeed not the case, then there would be little relevance to the possibility of *losing* the sense of a self in the process of attaining nirvana.

In Part Two, I argue that the process of attaining nirvana (as construed in Pali Buddhism) would place important constraints on the structure of the self-illusion, setting it apart from standard Western accounts of no-self. It is fairly uncontroversial that each person seems somewhat aware of a stream of various objects, including thoughts and perceptions, and that these objects seem presented, in our conscious awareness, to a point of view. This point of view appears to be unified, both synchronically and, from moment to waking moment, diachronically. (I term this point of view a ‘perspectival owner’.) But does the perspectival owner really exist? On most Western theories of no-self (pioneered by Hume), this unified conscious perspective lies at the heart of the self-illusion. All the conscious mind is really furnished with, it is claimed, are bundles of evanescent mental phenomena: thoughts and perceptions (etc.) that interact with mental mechanisms to create the illusion of a unified perspective that observes them. Of crucial relevance here is that the Buddhist theory of no-self is typically depicted by Buddhist scholars of different traditions as a type of bundle theory. While I have argued elsewhere that a bundle theory of no-self is not supported by specific sutras in the Pali Canon, the purposes of this paper are chiefly philosophical. To this end I offer a new argument against interpreting Buddhist sutras as bundle-theory reductionism—whether the reductionism is non-reflexive (discussed in this volume by Mark Siderits) or reflexive (discussed in this volume by Georges Dreyfus, Evan Thompson, Matthew MacKenzie, and Joel Krueger). I argue that if we take seriously the soteriological Buddhist injunction that the truth of no-self in nirvana is to be known experientially (e.g. through meditation) rather than just inferentially (e.g. through philosophical analysis), then a bundled mind will not be the sort of thing that can be known in the right way. The heart of the self-illusion will instead, I contend, lie in the personalized *identity* that seems to place a boundary around the (real) unified perspective, turning it into what I call a ‘personal owner’. I contend that this boundary, underpinning the sense of ‘who I am’ versus ‘what I am not’, is the true target of early Buddhist practices that seek to eliminate the sense of a self. What remains after the

sense of self has dissolved is a unified perspectival ‘witness-consciousness’ that, insofar as it lacks the illusion of a personal self, is intrinsically ownerless.

If nirvana is possible, and my arguments are accepted, then it has implications for how the everyday mind is structured. The personal-owner-self will be an illusion, while the unified witness-consciousness, which comes through in the ordinary sense of self, will be real. The mind will therefore exemplify a ‘two-tiered’ rather than ‘bundled’ illusion of self. In Part Three, I offer some independent reasons for supposing that the two-tiered illusion of self is exemplified in the everyday mind.

2. The Central Notion of Self

If nirvana is the seeing-through of the self-illusion, then investigation into its possibility must be clear on the kind of self that is supposedly seen through. This is especially important if, as Dan Zahavi and Joel Krueger claim, definitions of the self abound (Zahavi, this volume, p. 66, Krueger, this volume, p. 34). It is also important that the ‘self’ in this case depicts something central, something that most of us have a sense of being. Should it turn out, for instance, that the ‘self’ of Buddhist sutras refers primarily to an immortal soul, the prospect of seeing through the illusion of self would be irrelevant to most people, who do not presuppose its existence.² Elsewhere, I spend some time extracting from early Buddhist sutras the relevant notion of the self, and then arguing that this notion of the self, which closely matches that alluded to by many Western philosophers (such as Hume, James, and Dennett), is presupposed in our modes of thinking and living.³ For reasons of space I will not recite the arguments here, although I will insist, *contra* Zahavi, that it is very much a notion ‘in line with our pre-philosophical, everyday understanding of who we are’ (Zahavi, this volume, p. 66). The purpose of this section is, thus, to elucidate what I take to be the central features of this assumed self, drawing in particular on a distinction that I make between perspectival and personal ownership. As already hinted, the distinction will figure centrally in my account of how the self is illusory and what survives its destruction. In essence, then this commonly assumed self is a

² This is not to deny that the Buddha did caution Brahmanical thinkers against becoming enamoured with more theoretical elaborations of the self, involving eternal, non-physical impartite entities serving as the vehicle of rebirth.

³ See Albahari (2006, chs. 2–4).

unified, unbrokenly persisting subject of experience with personalized boundaries and a perspective on the world. It is a thinker, owner, and agent that stands behind, and is somewhat in charge of, the stream of thoughts and experiences, as opposed to being constructed by them.

First and foremost, the self (that we have a sense of being) is a *subject*, as opposed to *object* of experience. ‘The subject’ describes that aspect of the ordinary self which is the inner locus of the first-person perspective: the conscious embodied viewpoint from which the world is apprehended. The subject's *modus operandi* is simply to observe or *witness* objects through a variety of perceptual and cognitive modalities. I hence use the term ‘witness-consciousness’ to describe the purely observational component that is common to all modes of conscious apprehension, perceptual or cognitive.⁴ While this term has sometimes been used by scholars of Advaita Vēdānta with additional metaphysical commitment (e.g. Bina Gupta 1998), or to convey a relation of dependence with experiential objects (e.g. Wolfgang Fasching, this volume), my use of the term is intended to be neutral on this front. That said, my usage is quite congruent with Fasching's, insofar as he writes that witness-consciousness is ‘nothing but seeing itself’ (as opposed to a thing that ‘sees’) and is experientially present to our conscious life (Fasching, this volume, p. 194). (Later I will argue, in agreement with Fasching, that witness-consciousness is unified and to some extent unbroken, but importantly, this is not built into my definition of the term).

I use the term ‘object’ to describe anything that can possibly be attended to by a (witnessing) subject: thoughts, perceptions, trees, bodies, actions, or events. Any conscious creature is uniquely positioned to observe (via witness-consciousness) an array of such objects as pains, thoughts, or its own body, from a perspective to which no other creature has direct access. Insofar as various objects appear to a subject's perspective, in this direct first-personal way, the subject can be termed *aperspectival owner* of the objects. While the perspectival owner or minimal subject is built into the self, it does not, as it stands, amount to a self: it is rather a mere locus for the first-person perspective. The relation that perspectivally owned objects bear to the subject *qua* perspectival owner matches what Dan Zahavi calls the ‘first-

⁴ For a full definition and defense of the reality of witness-consciousness, see Albahari (2009). In Albahari (2006), I specifically relate the notion of witness-consciousness to that of the consciousness aggregate (Pali: *khandhā*; Sanskrit: *skandha*)—one of the five conditioned elements that Pali Buddhism claims to constitute a person.

personal givenness' or 'for-me-ness' of experience, and it is at play whenever one speaks of 'my headache', 'my body', or 'my actions'. I don't concur with Zahavi, however, in holding that for-me-ness *is* a minimal self or subject, as for-me-ness is a dimension of the stream of experiences rather than the subject experiencing the stream. (This will become relevant later in the discussion).⁵

Of note is that the subject (and the wider self) cannot appear directly to itself in the focal manner of a perspectively owned object, so although seeming to have a subtle phenomenal character (and hence a *sense* of itself), which can be enhanced during meditational practice, the subject is perpetually elusive to its own focally attentive purview. *Elusiveness* is thus a key attribute of the minimal subject, *contra* Strawson (this volume) who insists that a subject can be attentively (or 'thetically') aware of itself as it is in the present moment.

The self, as I've indicated, is more than just an elusive subject *qua* perspective on the world: it is a subject assimilated with several other roles. First and foremost is the role of *personal owner*. A personal owner, a *me*, is a subject with an *identity* (or 'who-I-am-ness') as opposed to a merely impersonal point of view. In relation to this identity as personal owner, the existing ownership of various items, whether perspectival (such as thoughts or feelings) or possessive (such as houses or cars), also takes on a personal 'mineness' dimension. In relation to one's felt identity as a Michael Jackson fan, for example, an autographed record by the artist is felt to be personally, not just possessively owned. In this manner such items become warmly infused with a sense of mineness that goes beyond a rational recognition of one's legal status as owner of the object. Conversely, the sense of personal ownership or mineness towards an object will seem to reveal a facet of one's fixed self-identity as its personal owner, enhancing the sense of *me-ness*. A felt identity is made perspicuous through *identification*, where certain ideas (such as gender, race, character traits, basic roles) are appropriated to a subject's perspective, such that the world seems approached through their filter. As J. David Velleman puts it, identification occurs when a part of the personality 'presents a reflexive aspect to [one's] thought' such that it becomes one's 'mental standpoint' (Velleman 2002: 114). The most basic and pervasive role to be identified with is simply that of the perspectival

⁵ See, for example, Zahavi (2005a: 9–10) and Zahavi (2005b: 122–123). The distinction between witness-consciousness and for-me-ness is discussed further in Albahari (2009: 67–68).

owner. In this manner, the subject does not merely approach the world and its objects from an impersonal psycho-physical point of view (through whatever sense modality); it deeply *identifies* with that viewpoint as a concrete place where *I, the self*, am coming from. In tandem with perspectival ownership, then, the sense of personal ownership is almost always in operation whenever one alludes to such things as ‘my thoughts’, ‘my headache’, ‘my body’, or ‘my actions’.

While feelings of desire and attachment are possibly the most salient phenomenal indicators of a sense of personal ownership (and hence identity as a personal owner), these feelings need not be present for the sense of such ownership to exist. This is particularly true in cases of profuse and mundane phenomena such as thoughts, sensations, and one's body. The sheer ubiquity with which such items are presented as personally ‘mine’ makes it quite impossible, in normal cases, to discern the distinct phenomenal quality of such ‘mineness’. For, as it happens, most people do not know what it is like to lack it. It is mainly pathological impairments, where the sense of personal ownership seems lost or compromised, that draw attention to the fact that there is this other major type (or sense) of ownership alongside the perspectival and possessive varieties. Subjects of anosognosia, for instance, may feel that a paralysed limb ‘does not belong’ to them, while subjects of depersonalization commonly sense a disconnection in ownership from many of their thoughts.⁶ While in cases of depersonalization the lack of personal-ownership feelings tends not to be global (e.g. they still identify as the subject of the dreadful condition that has befallen *them*), there is nevertheless a notable lack of identification with the perspectival owner of those thoughts from which they do feel disconnected. With reference to any object, then, a lack of personal mineness goes in tandem with a lack of personal me-ness. (Note that a possible suspension in the sense of personal ownership does not by itself prove such ownership to lack independent reality, any more than its suspension during deep sleep would prove its lack of reality. This point will be returned to later).

Subsumed under the role ‘personal owner’ are other frequent modes of identity that further delineate the type of self we take ourselves to be. Two

⁶ A level of disconnection in the sense of personal ownership may sometimes occur via the mode of agency, such that one feels that someone else has authored a particular line of thought, such as during ‘thought insertion’ in schizophrenia. As I consider the sense of agency to be grounded in a sense of personal ownership, the general point remains: a compromise in identification as author of the thought is a compromise in some level of identification as its personal owner.

closely related such modes are *agent* and *thinker*. ‘The agent’ is the owner-subject in its capacity of initiating actions. Taking pride or being ashamed of perspectively owned actions is an obvious way of identifying with the perspectival owner *qua* agent, such that one deeply feels ‘I am the initiator of this action’ (think of the proud winner of an Olympic medal). Such emotions also provide evidence of regarding oneself to have special causal powers that enable the active choice of one course of action over another, as opposed to a passive determination by the flow of events. To feel guilty, for instance, implies an assumption that one should not have acted in a particular way—and hence, arguably, an assumption that one could have acted differently. Intentional actions originate in thought, so the *thinker* is a closely related mode of identity. Importantly, we take ourselves to be the originator, controller, or observer of the thoughts, rather than to be, in essence, the content of thought. Put in terms of an attribute, the assumed self is something that is *unconstructed*—that is, not constructed from the content of thoughts and perceptions, etc.—some underlying thing that is their precedent rather than their product.

Any sense of identity, whether with a general role such as owner or agent, or a specific idea about who one is (for example, a female ice-dancing champion), evokes the elusive feeling of being bounded by that identity. Faced with the world and its objects, the subject thus reflexively presents not merely as a point of view, but as a *unique and bounded thing* with a point of view. This attribute of *boundedness* is absolutely central to the assumed self: it turns the perspectival into a personal owner. The bounded self seems, moreover, to be perfectly *unified*, in that its differing and shifting roles and identities (such as personal owner/actor/thinker, female/skater/champion) appear seamlessly integrated within the very same subject. A feeling of excitement at the upcoming ceremony may simultaneously trigger all the different roles, but it does not feel to the subject as if each role or identity corresponds to a numerically distinct self, or even to different compartments within a single self. Importantly, the field of unity seems to extend beyond the roles of the subject to share in the set of objects perspectively and personally owned by it. Perceptions, thoughts, and experiences, felt as belonging to the subject, present as belonging to the subject's very same field of consciousness, such that it seems natural to say: ‘I, the self, am simultaneously aware of the white ice and the cheering crowd’, or ‘The very same self that a few minutes ago saw the white ice heard the cheering crowd’. While the unity of

consciousness is a philosophical topic unto itself, the point here is simply that *unity* is a feature central to the commonly assumed self at stake in this discussion.

Unity is not only synchronic but diachronic. *Unbrokenness* describes that aspect whereby unity of the self is diachronically extended, and it is useful to divide this into (i) the specious present (temporally ordered over a very short interval), (ii) the span from one specious present to the next, and (iii) the longer term, whether awakening from deep sleep or persisting over a lifetime. *Invariability* captures the phenomenal side to this assumed identity—the elusive feel of it being the very same underlying ‘me’ that belies all kind of change to body and personality. So in addition to the flux of experience, famously noted by Hume, there is the sense of a background unbroken *me* that observes the experiences, and that typically persists beyond the scope of a waking episode. For example one may, upon awakening from deep slumber, flinchingly recall an embarrassing venture from the night before—or perhaps four years before—indicating a strong implicit identification with the perspectival owner of the regrettable action. Or one may awaken with a brilliant plan for the future. Such identification as the ‘longitudinal’ self helps solidify the sense of being a personal owner, such that the self takes on what Antonio Damasio (1999) calls an ‘autobiographical’ dimension (although this feature does not seem *essential* to the self).⁷

I have mentioned *unconstructedness* in the context whereby the self is tacitly but deeply assumed to be that which underlies and originates the thoughts, rather than that which is constructed by them: their precedent rather than their product. This assumed feature actually underpins the entire self. For as the self *qua* thinker presents implicitly as that elusive, unbroken, bounded subject which is *unified* with the roles of agent, owner, experiencer, and observer, the feature of unconstructedness will extend naturally to each and every role and attribute of the self. So if the self truly exists, as per its manner of presentation, then every one of its features will be unconstructed—the precedent, not the product of the stream of objects (thoughts, experiences, perceptions) to which it tacitly seems opposed.

⁷ Galen Strawson defines an ‘Episodic’ person as such: ‘one doesn’t figure oneself, the self or person one now experiences oneself to be, as something that was there in the (further) past and will be there in the (further) future’ (Strawson 2008: 210). He contrasts this with ‘Diachronic’ persons who do have a sense of identity with their earlier and future selves.

3. How Nirvana, as Construed in Pali Buddhism, Structures the Illusion of Self

How, in general terms, might this self turn out to be illusory? Could it be illusory in more than one way? These questions are addressed in this section, where I propose that the standard Western ‘bundle theory’ of no-self differs markedly from how no-self ought to be understood if nirvana, as depicted in Pali Buddhism, is possible. As already mentioned, this will involve a major challenge to the typical forms of Buddhist reductionism that cast no-self in Buddhism as a type of bundle theory.

What does it take, in general terms, for something to be illusory? Illusions (including delusions and hallucinations) essentially involve a mismatch between appearance and reality, such that something appears to be a particular way, when, in reality, it is not actually that way. Typical cases include the Muller Lyer Illusion (two lines appearing to be of unequal lengths when they are actually of equal length), a hallucination of the pink elephant, the sense of being watched by extra-terrestrials. In all these instances, the world does not veridically underpin the way the world appears to be, whether the medium of appearance is perceptual or cognitive. If the self is illusory, then the world (which includes the world of subjectivity) must similarly fail to deliver at least one of its defining characteristics as presented via the self’s characteristic mode of appearance.

Given that the self purports to be something that is entirely unconstructed in all its defining features, a straightforward route to casting the self as illusory suggests itself: argue that at least one of these features is constructed from the content of those thoughts, experiences, and perceptions to which the self seems opposed. In this way, the self will not be what it fundamentally appears to be—it will be (at least in part) the product, rather than the precedent, of thoughts and experiences. Put another way, the *sense* of self, which presents as being thoroughly grounded in (and actually identical to) the (unconstructed) self, will be grounded in factors other than this self.

A word about the sense of self is in order here, since ‘self’ and ‘sense of self’ are sometimes confused. The sense of self is the *appearance* of a self, pertaining to the reflexive feeling or conscious impression of being a self. Throughout the discussion, I have been supposing that this feeling, although elusive to attentive purview, is real enough. What is in question is the veracity of its content: the self. Indeed, philosophers who deny the existence of the self

do not generally deny the sense of self, any more than those who deny the existence of libertarian free will would deny the common feeling of such free will. In fact, as libertarian free will is sometimes (although not always) ascribed to the self in its capacity of being an agent, it can, because of the history of debate on the subject, provide a useful illustration of just how it is that a feature of the self (that we assume we are) may fail to reflect reality: how the sense of self, in other words, may fail to be grounded in a self.

Suppose, then, that determinism is correct, leaving no ontological room for an entity that could genuinely originate one course of action over another (the past being what it is). Suppose also that we (most humans) have a sense of being a self with this controversial sort of agency (a feeling that may be evidenced through such emotions as guilt). The feeling of free will will reflexively convey the cognitive content of being an entity that really does exercise such agency—not just of appearing to exercise it. So if determinism is to rule out the reality of such agency, then the sense of being a self *qua* libertarian agent will not be grounded in an actual agent-self, as it appears to be, but (at least partially) in the *content* of thoughts and feelings to which the self seems ontologically opposed. The thoughts and feelings will, in other words, be helping to create the conscious impression of there being a source of agency that is able to exercise libertarian control over the thoughts and feelings and actions—when, in reality, there is no such source of agency. The self, *qua* libertarian agent, will thus be an illusion created by the content of thought (etc.). It will fail to exist in the essential manner that it purports to exist, as something that stands entirely behind the thoughts, authoring the intentional actions.

To recapitulate: if the self exists, then it is an entity with a conjunction of unconstructed essential roles and features. Should at least one of these roles or features (such as agency) turn out to be mentally constructed, then the self, so defined, will not exist. The usual (but unacknowledged) strategy at the heart of most attempts to deny the existence of self is thus to argue that at least one essential feature of the self is a mental construct. Seeing this strategy at work will help to determine more exactly how the typical Western construal of no-self could differ from the way in which the self would fail to exist if nirvana were possible. I first examine some standard Western accounts of no-self before comparing this analysis to the case at hand.

David Hume is commonly considered to have pioneered the Western philosophical position on no-self, his work sometimes compared by scholars

to the Buddhist no-self doctrine. He argues that, instead of there being an unbroken, underlying entity which unites the varying perceptions and accounts for their identity, as there appears to be, there is merely:

a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement.... There is properly no simplicity [unity] in it [the mind] at any one time, nor *identity* in different, whatever natural propension we may have to imagine that simplicity and identity. (Hume 1739: I, IV, vi)

As for what accounts for the *sense* of identity:

The identity which we ascribe to the mind of man is only a fictitious [viz., constructed] one, and of a like kind with that which we ascribe to vegetables and animal bodies. It cannot therefore have a different origin, but must proceed from a like operation in the imagination upon like objects.... identity [and simplicity] is nothing really belonging to these different perceptions, and uniting them together, but is merely a quality which we attribute to them, because of the union of their ideas in the imagination when we reflect upon them. (Hume 1739: I, IV, vi)

Hume regards the appearance of the self's unity and identity to be underpinned, not by factors that include *actual unconstructed* unity ('simplicity') and unbroken, unchanging identity ('uninterruptedness' and 'invariability'), but by mental factors such as 'the union of their ideas in the imagination when we reflect upon them'. The self, in other words, is a mental construct *by virtue of* the fact that unity and (short and long-term) identity have constructed status. The illusion is hence that of an unconstructed entity, a self, whose features of unity and unbroken identity objectively underpin our *sense* of unity and identity. In reality, there are no such principles of unity or identity that actually underpin this impression, either in the self or in the mind. There is only a diversity of rapidly fleeting perceptions that, when acted upon by the memory and imagination, create the impression of an entity with unity and identity.

Casting unity and unbroken persistence as central to the self's constructed, and hence, illusory status turns out to be a strategy common to virtually all Western accounts of no-self. For instance, William James makes it clear that

the unity and unbrokenness we commonly ascribe to the self are unconstructed:

common-sense insists that the unity of all the selves is not a mere appearance of similarity or continuity, ascertained after the fact. She is sure that it involves a real belonging to a real Owner [a source of unity], to a pure spiritual entity of some kind. Relation to this entity is what makes the self's constituents stick together as they do for thought.(James 1890: 337)

On James's position, the self's supposed unity and unbrokenness is, as Owen Flanagan puts it, 'an after-the-fact construction, not a before-the-fact condition for the possibility of experience' (Flanagan 1992: 177, 178). Flanagan's own defense of the no-self doctrine follows James' insofar as he bestows illusory status to the self principally via the features of unity and unbrokenness (with other features such as agency riding on this):

The illusion is that there are two things: on one side, a self, an ego, an 'I', that organizes experience, originates action, and accounts for our unchanging identity as persons and, on the other side, the stream of experience. If this view is misleading, what is the better view? The better view is that what there is, and all there is, is the stream of experience. 'Preposterous! What then does the thinking?' comes the response. The answer is that 'the thoughts themselves are the thinkers' (James, 1892, 83)⁸...We are egoless.(Flanagan 1992: 178)

On a similar theme, Daniel Dennett writes:

Each normal individual of this species makes a *self*. Out of its brain it spins a web of words and deeds and, like other creatures, it doesn't have to know what it's doing; it just does it...Our tales are spun, but for the most part we don't spin them; they spin us. Our human consciousness, and our narrative selfhood, is their product, not their source...These strings or streams of narrative issue forth *as if* from a single source...their effect on

⁸ Flanagan cites William James (1892), *Psychology: The Briefer Course*, G. Allport (ed), New York: Harper and Row, 1961.

any audience is to encourage them to (try to) posit a unified agent whose words they are, about whom they are: in short, to posit a *center of narrative gravity*. (Dennett 1991: 418)

These thinkers typify the way in which the self is denied in Western philosophy, by giving constructed and thereby illusory status to the central unified subject, where the unity is understood to be both synchronic and diachronic. While the accounts differ in their details of how exactly the impression of unity and unbrokenness is constructed—such as which mental faculties contribute to the illusion—all of them deny the existence of self principally via this avenue. The impression of unity and unbrokenness, as it qualifies a minimal subject (or perspectival owner) standing opposed to the stream of experience, must be entirely fabricated from the bundle of discrete mental phenomena to which it seems opposed. Essentially, they are what are known as *bundle theories* of the self.

I now ask: is this way of understanding the self as illusory—as a bundle theory that denies unconstructed reality to unity and unbrokenness to the self *qua* minimal subject—in line with how we should understand the status of ‘no-self’ if nirvana, as depicted in the Pali sutras, is to be possible? As I noted in the Introduction, this is exactly how Buddhist philosophical tradition has typically understood the doctrine of no-self. To this end, there are different versions of the bundle theory. The more extreme version (a non-reflexive reductionism inspired by Abhidharma tradition), has it that all impressions of unity—synchronic and diachronic—are illusory constructs from an ontology of momentary, causally connected aggregates. A less extreme version of bundle theory (which I call ‘reflexive reductionism’, inspired mainly by Yogācāra-Sautrāntika tradition) allows a measure of synchronic unity to exist through each conscious experience being reflexively aware of itself (perhaps for the length of a specious present).⁹ Such temporal unity, however, does not extend beyond the specious present, and it belongs not to any subject, but to the discrete experiences that form the changing stream of consciousness. All variants of bundle theory within the gamut of Buddhist tradition thus uphold the unreality or illusory status of an unbroken and unified witness-consciousness which, as *modus operandi* of the (minimal) subject, stands apart

⁹ My use of the term ‘reflexive reductionism’ is intended to depict the allegiance to bundle-theory, rather than reflect how its advocates would label their position (they would probably not, in their own context of use, call it ‘reductionist’).

from and observes the stream of experience. (Note that while Dreyfus (this volume) does not regard reflexive reductionism as a bundle theory, it counts as a bundle theory for the purposes of this discussion). Now I am well aware that the argument about to be offered, which defends nirvana (in the sutras) as entailing the unbroken unity of observational witness-consciousness (at least during the scope of waking life), flies in the face of many Buddhist philosophical traditions—and so will in that sense be denied by many to be truly ‘Buddhist’. So be it. What I do contend is that my position offers a more coherent philosophical reconstruction of the early Buddhist sutras than the bundle theory, and that the position, although not stated explicitly in the sutras, is quite consistent with them.

Before commencing with the argument, I need to say more about how reflexive and non-reflexive varieties of reductionism are to be distinguished. According to reflexive reductionism (discussed in various forms by Dreyfus, Thompson, MacKenzie, and Krueger in this volume) ‘consciousness cognizes itself in cognizing its object’ (Siderits, this volume, p. 318), so there is nothing more to consciousness than the cognizing experiences themselves. Put another way, the immediate object of a cognition is not the object out there in the world (such as the blueberry): it is the phenomenal experience of cognizing the object and ‘the experience of seeing blue is just the occurrence of a cognition that has blue color as its form’ (Siderits, this volume, p. 317). A stream of different consciousnesses thus amounts to a stream of multi-modal experiences: there is no separate cognizing subject. According to non-reflexive reductionism (discussed by Siderits in this volume), consciousness is not self-intimating in this way: it is an object-directed awareness that arises in conjunction with the various sensory or mental objects (including experiences) that form its intentional content. I address each version in turn, beginning with the latter.

On Buddhist non-reflexive reductionism ‘consciousness arises in dependence on contact between sense faculty and sensible object...the consciousness that takes the color of the flower as object must be distinct from that which takes its smell as object a moment later’ (Siderits, this volume, p. 313). Each momentary consciousness imparts information to the next moment: there is no temporal gap between each moment. Each composite moment of object-directed consciousness (if it is, *contra* reflexive reductionism, more than just the occurrence of the target mental or sensory object) will presumably have built into it an invariant observational

component, by virtue of which it is labeled ‘conscious’, and by virtue of which the illusion of unbroken observational consciousness, central to the illusion of self, is generated (just as the illusion of a continuous unified circle of fire in a whirling firebrand is generated by the invariant fiery nature of each distinct occurrence of the fire). Now I think that despite their differences, Buddhist traditions would converge in supposing that nirvana involves a transformative insight into the nature of mental reality that is based primarily upon first-person experiential observation, rather than intellectual puzzle-solving. One cannot get enlightened simply by studying philosophy or calculating laboratory results. On the Buddhist (non-reflexive) reductionist picture, then, the ‘person’ who has attained nirvana will directly ‘see through’ the illusion of self by viewing the nature of persons to be nothing but a fleeting causally connected bundle of discrete aggregates—including the very cognition that discerns it to be so. Our previous deep-seated assumption of being an unbroken, unified conscious self that witnesses the flow of phenomena will be dramatically overturned, as that consciousness itself becomes discerned as part of the very flow. Burmese meditation master Venerable Mahāsi Sayādaw describes the keen level of discernment in a way that lends support to the reductionist picture:

And the dissolution of consciousness noticing those bodily processes is apparent to him along with the dissolution of the bodily processes. Also while he is noticing other bodily and mental processes, their dissolution, too, will be apparent to him in the same manner. Consequently, the knowledge will come to him that whatever part of the whole body is noticed, that object ceases first, and after it the consciousness engaged in noticing that object follows in its wake. From that the meditator will understand very clearly in the case of each successive pair the dissolution of any object whatsoever and the dissolution of the consciousness noticing that very object. (It should be borne in mind that this refers only to understanding arrived at through direct experience by one engaged in noticing only; it is not an opinion derived from mere reasoning). (Sayādaw, 1994: 23)

Given that nirvana must entail, as Venerable Sayādaw intimates, ‘an understanding arrived at through direct experience’ as opposed to ‘an opinion derived from mere reasoning’, I think that there is something incoherent about

a picture which, when taken literally, has the meditator experientially aware that their discerning consciousness is impermanent. I contend that the best way to make sense of such a passage is to suppose that the meditator is actually aware of different *directions* that are being taken by consciousness (in virtue of the objects)—a picture that does not entail Buddhist reductionism. So when the Pali sutras speak of consciousness as being impermanent, I take this to mean that the intentional content of consciousness—that to which consciousness is directed—is constantly changing. One moment there is consciousness of green and round, and the next, consciousness of crunching and apple-taste. But this is not the same as saying that the observational component that is directed towards these objects is itself arising and passing away. So what is my argument against this interpretation of Pali Buddhism?

Let me first be clear on what I am not arguing. I am not arguing that the impression of unified unbroken consciousness is impossible under reductionist ontology. My argument is instead based upon an epistemic aspect that grows out of the idea that the primary mode for understanding the mind, in nirvana, is experiential. Nirvana is often depicted as ‘ultimate’ in the Pali sutras, not only in an axiological but an epistemic sense. Statements made from the nirvanic perspective are taken to be authoritative: there is never the idea that they could be usurped by philosophical or scientific discovery. For example, there is never any intimation in the Pali sutras that the Buddha or arahant could be mistaken in saying such things as ‘conditioned phenomena are impermanent, conducive to suffering and without a self’. I call this the ‘Experience Condition’:

Experience Condition: The primary mode of knowledge/wisdom/insight, in the nirvanic state, is based on first-person experience and the first-person perspective has authority over the third-person theoretical perspective. In cases of a conflict between first-personal nirvanic perspective and third-person theoretical perspective, regarding the nature of the conscious mind, the first-person nirvanic perspective trumps the theoretical.¹⁰

¹⁰ Note that I am not defending the Experience Condition as a stand-alone condition in this paper, although towards the end of the paper it re-emerges as a suggested way to save the appearances. In relation to the concept of nirvana across Buddhist traditions, however, the Experience Condition may need further defense. For example, Siderits (this volume) points out that in Yogācāra subjective idealism, consciousness is regarded as ultimately non-dual in nature. Nirvanic realization will involve a complete dissolution of the idea that there is an external world with its objects—and hence, according to Siderits, a dissolution in the feeling of subjective interiority which must depend upon the subject/object split. Thus ‘Yogācāra subjective idealism involves an explicit disavowal of the perspectival self argued for by Albahari’ (this volume p. 318). So if the ideal nirvanic state involves no sense of interiority or first-person perspective, then how can

I will argue that the Experience Condition entails (on the nirvanic hypothesis) that the aspect of consciousness which cognizes the impermanence of phenomena is neither discrete, nor reducible to the changing stream—at least during wakefulness. I begin by asking: how would the cognizance of an impermanent consciousness (by an impermanent consciousness) work on a reductionist account? It could not be that each moment of consciousness reflexively observes its own coming and going. For regardless of whether or not we are dealing with reflexive consciousness, it would entail the contradictory state of a conscious moment being present to its own coming and going. The impermanence (or diachronic disunity) of the discerning cognitions must therefore, on any reductionist account, be experienced retrospectively.¹¹ So let us suppose that a discerning cognition at t_3 is a momentary member in a causal chain of conscious moments t_1 - t_n and that it retrospectively (whether via memory or retention) discerns the impermanence of prior members of the chain. As this cannot include itself (t_3), it will have to discern, say, the numerical transition from moments t_1 and t_2 , before itself becoming retrospectively discerned as impermanent by a later cognition, say t_5 . And here is where I see the problem for *non-reflexive* reductionism, for the only way that t_3 can experientially distinguish the transition from t_1 to t_2 is by discerning the changing *content* of t_1 to t_2 —the *object* towards which each consciousness is directed. But this does not tell us that at each moment, the observational component is in fact numerically discrete—it just tells us that the objects are. There is in fact no way to phenomenologically tell whether the underlying ‘objective’ scenario (if there is a further truth to the matter) is that of a contiguous chain of numerically

nirvanic authority be indexed to a first-person perspective? Here we must tread carefully; ‘first-person perspective’ or ‘interiority’ is ambiguous. If it means ‘experience confined to a dualistic subject/object structure’ then I would agree that ultimate non-dual consciousness must lack this structure, and so the first-person perspective must lack ultimate authority. But if it means that ‘there is something it is like to experience non-dual consciousness’—and elsewhere (2006) I attempt to convey in some detail what this could mean—then ‘first-person perspective’ would not be disavowed by non-dual consciousness. (I suspect that it lies behind many of the intimations that nirvana is experienced, not inferred). Be that as it may, I would insist that so long as objects *are* experienced, the dualistic (subject/object) first-person perspective, with its perspectival subject, is unavoidable. And so long as the domain of judgment from the nirvanic perspective is about subjective experience in its relation to its objects, then the Experience Condition, even if construed narrowly, remains intact. If it cannot but seem as if objects are being witnessed by (a subject’s) unbroken conscious awareness, then (if nirvana is possible) this will indeed be how things are, even if the unbroken awareness, in its intrinsic nature, is not confined to the perspective of a subject.

¹¹ I put aside any problem that might arise with elusiveness of the discerning cognition, as it is not clear that non-reflexive reductionists would accept this aspect, and I want to engage in the debate on their terms.

discrete discerning consciousnesses, generating the illusion of unbroken consciousness taking different objects, or that of an unbroken discerning consciousness cognizing an array of different objects—in line with how things seem. That is because the observational component, which renders each moment of non-reflexive consciousness to be *conscious*, is qualitatively invariant, leaving no marker by which the contiguous numerical transition could be experientially discerned (it's not as if there will be a little jolt at each transition). The observational component to each conscious moment will thus seem, from the first person experiential perspective, to be unbroken—regardless of the underlying ontology. To use Thomas Metzinger's (2003) phrase, the mind will be 'phenomenally transparent' to the extent that any underlying ontology, regarding the impermanence of the discerning cognition, will be inaccessible to subjective experience. This is a problem for non-reflexive reductionism. We should expect the nirvanic insight into impermanence and no-self to be experientially and cognitively dramatic, invoking a shift from the incorrect to the correct perspective. But how could this happen, if there is no way to phenomenologically discern the incorrect (unbroken) from the correct (discrete) state?

At this point, the non-reflexive reductivist may reply: if the nature of the discerning cognition cannot be determined one way or the other, then we cannot conclude that the discerning cognition is *unbroken*, any more than we can conclude that it is discrete. Here is where the Experience Condition comes in. If it is accepted that the nirvanic first-person perspective is authoritative on the nature of our mental life, and that from the nirvanic perspective it cannot but seem as if there is unbroken observational consciousness from one moment to the next (sustained throughout waking life), then there won't *be* a hidden ontology at variance with the appearance. Things will be as they seem: appearance and reality will converge. Inferring the discerning consciousness to be discrete (as Siderits does) involves, *contra* the Experience Condition, taking the third-person perspective, with its method of philosophical inference, to be authoritative. In Venerable Sayādaw's terms, it is putting 'an opinion derived from mere reasoning' ahead of 'understanding arrived at through direct experience'. If the discerning consciousness does turn out to be discrete, it cannot therefore match the model of non-reflexive reductionism.

Does reflexive reductionism fare any better? The problem with non-reflexive reductionism arose because the invariant observational component within each moment of consciousness made it impossible to discern its

numerical transition from one conscious moment to the next. Reflexive reductionism does not appear to succumb to this problem, for any impression of unbroken and invariant conscious observation must crumble under sharper scrutiny. Why? Because there is no more to consciousness than simply the stream of experience itself (e.g. visual, auditory, proprioceptive)—there is no separate observational component that takes each experience as an object. Each experience is reflexively aware of itself, which is what makes it a *conscious* experience. And as each conscious experience is qualitatively different, it will be quite possible to retrospectively tell (say at t_3) when a moment of consciousness (say t_1) has ceased and another (t_2) has begun. Hence each reflexive moment of consciousness will be experientially discernible as impermanent, such that it does not fall prey to the Experience Condition. The findings from this first-person perspective of authority will trump any theory that insists, in line with the self-illusion, that there really is a minimal unified subject that unbrokenly observes the changing phenomena. Any impression of a separate invariant subject of experience will be generated through cognitive processes (and theoretical accretions) that have us paying insufficient attention to the degree of change in our conscious life. Nirvana will thus entail insight into the real fact that the degree of change, being far more dramatic than we'd assumed, does not support an unbroken self (or indeed subject) that observes the stream.

It may be tempting to object that while non-reflexive reductionism goes *beyond* the scope of experience, by attributing numerical discreteness to an invariant observational component of consciousness, reflexive reductionism *under-describes* the scope of experience, by failing to acknowledge that there is more to conscious experience than just the manifold. For does not the very idea of a nirvanic *first-person perspective* convey that there is more to our conscious life than *just* the manifold of experience, and might this dimension not point to an invariant observing subject? If reflexive reductionism fails to describe the phenomenology of conscious experience properly, by leaving out the first-person perspective, then it will, at the very least, not satisfy the Experience Condition.

This objection, as it turns out, misconstrues reflexive reductionism—or at least the versions that have been portrayed in certain schools of Indian, Chinese, and Tibetan Buddhist philosophy (see Thompson, Dreyfus, Siderits, MacKenzie, and Krueger, this volume, for a discussion of these positions). If

we look more closely at the kind of phenomenal considerations advanced in favor of reflexive reductionism, we will see that the first-person perspective is not ignored. For instance, in his discussion of the position (this volume), Dreyfus speaks of the ‘first-person self-giveness’ of experience. The notion of first-person giveness has been articulated at some length by Zahavi (2005, this volume), whose phenomenologically inspired position shares with reflexive reductionism the central tenet of bestowing reality to only the stream of experience, rather than to a witnessing subject of the stream (his insistence on diachronic unity between specious presents within the stream differentiates him from the full-fledged reflexive reductionist). Of first-person giveness, Zahavi maintains that our conscious experience is not deeply impersonal or anonymous but is structured by a fundamental self-giveness in such a way that ‘the experiences I am living through are given differently (but not necessarily better) to me than to anybody else’ (Zahavi, this volume, p. 59). Hence no matter how similar my experience of the green apple is to Miguel's, there is a perspectival belongingness between my experiences that make them cohere together as ‘mine’ in a way that they do not with any of Miguel's experiences. Zahavi continues: ‘...anyone who denies the for-me-ness or mineness of experience simply fails to recognize an essential constitutive aspect of experience. Such a denial would be tantamount to denial of the first-person perspective’ (Zahavi, this volume, p. 59). Or as Evan Thompson puts it, the experience (if from memory) ‘is given from within as an experience formerly lived through first-personally, that is, by me’ (Thompson, this volume, p. 173). In other words, all our experiences are presented through a perspective and this perspectival for-me-ness reflexively had by each one of our experiences constitutes our first-person perspective (as the experiences seem to be *for* the same *me*), lending our experiential life the *impression* of unity across time.

But while this may now correctly describe our experience, the type of problem that plagued non-reflexive reductionism threatens to re-emerge. For how can we phenomenally tell whether the situation is that of numerically discrete and contiguous (qualitatively invariant) for-me-nesses, or just one unbroken for-me-ness, which may well point to an invariant observer? The reflexive reductionist will again respond that this misunderstands their position—we *can* tell. Just as the identical sheen belonging to each colored bead on a string doesn't prevent us discerning the different beads by their color and shape, the identical reflexive for-me-ness belonging to each

conscious experience doesn't prevent us from discerning the different experiences by their varying qualities. For remember: on reflexive reductionism, a qualitative change within experience amounts to a numerical switch in consciousness; diachronic unity (at least beyond the scope of the specious present), along with the subject as locus for synchronic unity, is fundamentally an illusion.

But we may now wonder where this leaves the commonly asserted idea of an invariant and separate subject standing apart from and observing the stream of experiences. Are philosophers such as Hume, James, and Dennett uniformly mistaken in their description of the phenomenology of our experience—and indeed, of the very sense of self enumerated earlier in this piece? Is the minimal observant subject no more than a careless theoretical reification that evaporates under careful phenomenal scrutiny? It would appear so, according to its proponents. Reflexive reductionists are generally adamant that a separate perspectival observer (or observational component) is not to be introduced into either the phenomenology or the ontology of conscious life. In keeping with this aspect of the position, Zahavi goes so far as to redefine ‘the self’ as the subjectivity of experience: ‘the self is defined as the very [invariant] subjectivity of experience and is not to be taken to be something that exists independently of or in separation from the experiential flow’ (Zahavi, this volume, p. 60).

The crucial issue at this stage is not that of whether there really *is* an invariant observational component (or perspectival owner) that is intrinsic to our conscious life, but whether there *seems* to be—or more accurately, whether there *must* seem to be (such that it survives phenomenal scrutiny). While I agree that for-me-ness characterizes our experience, I contend that it structures our conscious life far more dramatically than as just a reflexive sheen on the bead of each experience. It necessarily *bifurcates* our experience into subject and object. For so long as our diverse experiences seem to be *for* me—and for the very same me over time, no less—there is no escaping that there will seem to be a perspectival ‘me’ that the experiences are for. Or to put it more simply: so long as objects are experienced as being given to a subject there must seem to be *a subject* to which they are given. (Galen Strawson, this volume, p. 275, argues for a logical corollary of this position, claiming that so long as there is subjectivity, there must, as a matter of logical fact, be a minimal subject—although I am not going so far as to claim yet that the subject is *real*.) This strongly suggests that the phenomenology of

experience contains more than simply the flow with its first-personal givenness. As a matter of phenomenal necessity, there will seem to be a perspectival owner that stands apart from the stream, such that it cognizes the experiences that are first-personally ‘given’ to it.¹²

Suppose the reflexive reductionists re-assess their phenomenological stance, agreeing that the ascription of first-person givenness entails the (elusive) appearance of a subject to which the experiences are given. They are still free to defend the *ontological* side of their position by insisting that, despite appearances, there is no separate subject of experience. If they are correct in this assessment, then the witnessing subject will be a mere illusion projected forth by the invariant dimension of an otherwise diverse stream of experiences. If a subject-realist is correct in their ontological assessment, then the appearance of the observing subject will reflect how things actually are. And here is where reflexive reductionism *does* get into trouble. Just as the dispute over whether the observational component is unbroken cannot be resolved without going beyond the first-personal appearance and appealing to philosophical analysis, the dispute over whether the minimal observing subject (to whom experiences are given) is chimerical cannot be settled without going beyond how things must appear. So if reflexive reductionists are correct in supposing the subject of experience to be chimerical rather than real, then nirvanic insight into this fact will have to be purely intellectual, rather than experientially based. With no experiential avenue through which to characterize the potential shift from incorrect to correct perspective, the dramatic nirvanic insight into no-self will be left unaccounted for.

But as mentioned earlier, the Pali sutras do not leave such matters unsettled: matters are arbitrated by the Experience Condition, which privileges the first-person perspective of nirvana. If it must seem, from the nirvanic perspective, that there is a minimal subject of our changing experience, then this will not be a mere appearance that can be usurped by theoretical inference. The limits of nirvanic appearance will dictate the scope of mental reality. So if nirvana (as depicted in the Pali sutras) is possible, reflexive reductionism does no better than the non-reflexive version at eliminating the invariant and unbrokenly observing subject of experience. The cognitive transformation of nirvana, I conclude, cannot be an insight into the fact that

¹² I develop this line of argument from another angle in Albahari (2009).

our mental life lacks such a perspectival owner and entirely comprises fleeting cognitions.

This leaves us with the question: in virtue of what features should we say that the self is constructed (hence illusory), if nirvana, as depicted in Pali Buddhism, is to be possible? How do we construe the cognitive transformation whereby the constructed status of the self is seen through? I contend that personal ownership and boundedness (and agency to the extent that it requires identity as a personal owner of the actions) are the most likely features of the self to be constructed. To reiterate: a personal owner is a perspectival owner that has identified with a variety of roles (including the basic role of perspectival owner), such that the bare witnessing perspective is cemented into a definitive thing with an identity (a 'me'). It is a subject with personalized *boundaries*, which personally (not just perspectivally) owns its thoughts, perceptions, feelings, experiences, and possessions. Given that there are known pathologies which compromise the usually ubiquitous sense of bounded identity, it is quite possible to conceive of a state, akin to global depersonalization, where all sense of bounded identity is lost. This opens up the distinct cognitive potential for a transformative experiential insight into the reality of no-self, although by all accounts it will not be pathological.

On this hypothesis, the illusion of self will arise through the mechanism of identification. Identification (to reiterate) is the appropriation of mental content to the subject's perspective, such that the content seems to qualify (and hence filter) the very outlook through which the world is approached. To the untrained perspective, it will appear as if identification is not constructing, but revealing various aspects of the self's permanent, prior existence. But the bounded self will in fact, on this hypothesis, be constructed through the process of identification. On the face of it, this sounds rather similar to how MacKenzie describes the process of 'self-appropriation' (MacKenzie, this volume p. 264). There is a crucial difference, however. On MacKenzie's construal of Buddhism (a version of reflexive reductionism), the minimal unbroken subject *emerges* from the act of appropriation, giving it a constructed status (à la bundle theory), whereas on my construal of Buddhism, the minimal unbroken subject is the unconstructed locus of appropriation. With each act of identification, the perspectival owner imports its unconstructed, unified and unbroken witness-consciousness into the illusion of self, such that it appears that the bounded self is the originator of these qualities. And with each act of identification, the discrete mental content

(identified with by the minimal subject) colors the perspectival outlook of the subject, such that the unconstructed subject appears as a bounded personal owner. (The dual constructed/unconstructed contribution is the reason I call it the *two-tiered* illusion of self). To the extent that unity and unbrokenness are ascribed to a subject that seems personally bounded, the unconstructed features will themselves undergo a measure of distortion. Hence: thoughts, feelings, perceptions, and experiences will seem presented to a personal unified owner insofar as there is the sense of being a *someone* with a personal boundary that operates from, and is in charge of, the unified perspective. The personal unified owner/agent/thinker will assume a thicker diachronic identity as the natural (moment-to-moment) *unbrokenness* of witness-consciousness becomes folded into the impression of a bounded self with a life-history and anticipated future that plans, remembers, deliberates, and wishes.

In view of this analysis, I propose that nirvana, as a deep and transformative insight into no-self, be understood as the culmination of a process whereby the trained use of witness-consciousness, through meditation, brings about a full de-identification from all mental and physical phenomena.¹³ The result will be the undoing of the self illusion. How exactly the process of de-identification could work is a topic for further research, but the general idea, I contend, is as follows. Identification is the appropriation of highly impermanent mental content (objects) to a subject's first-person perspective. So long as the objects remain appropriated to the subject's perspective, as part of the self's 'unconstructed' identity, their status as impermanent objects will be effectively rendered invisible, such that the subject is change-blind to their coming and going. The process of meditation will train the subject's attention to become increasingly percipient of the degree to which mental phenomena do change. So long as objects are being viewed for what they really are—as changeable objects—they cannot be simultaneously appropriated to the subject's perspective. By extrapolation, a full observation of all perspectival objects in their true state of transience, from moment to conscious moment, will imply a complete lack of identification with any of them. I anticipate that the process of de-identification would gain extra momentum as the subject repeatedly observes,

¹³ The general strategy of attaining nirvana, through de-identifying from all phenomena, has support from the famous *Anattā-lakkhaṇa Sutta* in the Pali Canon. In this, the Buddha urges his disciples to lose the sense of personal ownership and identification towards all categories of object (mental and physical), such that there is the discernment: 'This is not mine. This is not my self. This is not what I am'.

with increasing clarity, the mechanism by which identification works (via its undoing). Just as uncovering a magician's trick makes it impossible to keep on being fooled by it, lifting the veil of identification will make it impossible to remain fooled by its content of self-identity. Viewing the real nature of consciousness as unbounded by ties of identification is, I hold, what nirvanic insight actually entails. It may be tempting to hold that in view of such a realization, the arahant now believes 'I am not a bounded self—I am witness-consciousness'. This is misguided, for such a belief would incur further identification and hence a new binding identity. Nevertheless, I contend that nirvana involves a direct realization that consciousness, in the impersonal sense, is ownerless, in the personal sense.

4. Implications for the Everyday Mind

At the outset, I intimated that investigation into the psychological possibility of nirvana could have important implications for how we are to understand the architecture of the everyday mind. I have defended a position on how to best interpret the doctrine of no-self if nirvana, as portrayed in Pali Buddhism, is indeed possible. So if nirvana *is* psychologically possible, it will place significant constraints upon the structure of the everyday mind, such that at the very least it (a) harbors an illusion of self that is (b) constructed in the prescribed 'two-tiered' manner (with the constructed and unconstructed contributors). (Conversely, should it turn out that there is a self, or that the bundle theory of no-self is correct, then nirvana, as I've depicted it, will not be psychologically possible). In this section, I sketch some independent reasons to suppose that the self is a two-tiered illusion. While this will not, of course, suffice to prove the psychological possibility of such nirvana, it will strengthen the case for supposing that it is possible. I approach this task in two parts: first, through gesturing at some empirical evidence that suggests the mind to have a structure *compatible* with a two-tiered illusion of self, and second, via an argument which suggests the structure to be indeed *best explained* by the two-tiered illusion of self (with constructed and unconstructed contributors).

Is there reason to suppose that the mind could be structured in a way that is compatible with the two-tiered illusion of self? What sort of empirical evidence might count towards such a hypothesis? I contend that the right sort of evidence would involve established cases, in neuropsychological literature,

where there appears to be the presence of perspectival ownership (and hence witness-consciousness) without any sense of personal ownership. For this combination, after all, is what survives in the ‘ownerless’ consciousness of nirvana. The potential absence of personal-ownership feelings would provide some evidence that personal ownership is not essential to conscious life, increasing the likelihood of it being constructed (and hence, amenable to deconstruction). Conversely, the perpetual persistence of perspectival ownership to conscious life would increase the likelihood that it is an essential, unconstructed feature of the mind (that can perhaps be utilized in an attempt to undermine the sense of personal ownership).

It is perhaps ironic, then, that the clearest evidence (outside the Buddhist domain) for perspectival ownership sans personal ownership lies in that of pathological brainstates—quite opposite to exalted nirvana. Earlier in the paper, I alluded to the pathologies of depersonalization and anosognosia, where the impairment appeared to compromise a degree of personal but not perspectival ownership over one's thoughts, feelings, and body. For example, in episodes of depersonalization, there is often a reported sense of disconnection from streams of perspectively owned thought, such that it is not uncommon to hear phrases like ‘the thoughts that cross my mind do not belong to me’. In these cases, it seems that there is little doubt that the type of ownership compromised is personal and not perspectival. However, the very fact that such subjects commonly report being distressed about their condition would suggest they still identify as the subject of the distressing symptoms and hence, as a personal owner that takes itself to be bounded by its predicament. What is needed is evidence of perspectival ownership coupled with a *complete* lack of personal ownership feelings.¹⁴

Such evidence, I contend, may well be found in the pathological impairments of epileptic automatism, akinetic mutism, and the advanced onset of Alzheimer's disease (Damasio 1999: 98), as well as in infants or perhaps people awakening from general anaesthetic. In an episode of epileptic or absence automatism, which can be brought on by a brain seizure, patients will suddenly freeze whatever they are doing, and after a few seconds of suspended animation, perform simple actions (such as walking about) with a

¹⁴ It may be suggested that creative absorption and highly attentive ‘flow states’ provide better examples of states where [the sense of] personal ownership has dropped away while perspectival ownership is present. Perhaps this is true. But, it is not apparent to me whether such states, short of being had by someone who is an arahant (hence presupposing the psychological possibility of nirvana), provide clear-cut documented evidence of cases where there is *no* remnant [sense] of personal ownership.

completely blank expression. Upon recovering from the episode, the bewildered patient will have no memory of what just occurred. Damasio, who had such patients in his care, writes:

[The patient] would have remained awake and attentive enough to process the object that came next into his perceptual purview, but inasmuch as we can deduce from the situation, that is all that would go on in the mind. There would have been no plan, no forethought, no sense of an individual organism wishing, wanting, considering, believing. There would have been no sense of self, no identifiable person with a past and an anticipated future—specifically, no core self and no autobiographical self.(Damasio 1999: 98).

This passage, I believe, provides good evidence to suppose that, through suspension of the ‘core self’ (corresponding closely to the *sense of self* that has been described in this paper), the feeling of personal ownership—with its trappings of identification and emotional concern—can be entirely absent, while bare perspectival ownership remains through the patient's simply being awake enough to respond minimally to his environment.¹⁵ (The ‘autobiographical self’ is an extension of the core self, by which the idea of one's personal history and anticipated future becomes integrated with one's assumed identity).

This case (and others) demonstrates the real possibility of a mindscape that is quite compatible with the model of the two-tiered self-illusion. But it does not in itself show the two-tiered model to *best explain* the phenomenon. There are at least two rival explanations that need ruling out. First, it could be that the self really does exist as an unconstructed feature of the mind, and that what has happened in such cases is that the *sense* of self has become suspended, while the self has persisted, holus bolus, all along. Not even the most ardent self-realists would wish to deny that the *sense* of self gets suspended during episodes of deep sleep. They would claim that during these unconscious phases, the self unbrokenly persists, such that upon awakening one *rightly* assumes it to have been the very same self that went to sleep. So why couldn't the self persist during the epileptic automatisms? Second, even if

¹⁵ Damasio (1999) himself seems to rule out linguistically the possibility of residual conscious awareness by routinely defining ‘consciousness’ as ‘consciousness with a sense of self’. For obvious reasons, I resist this move.

the self is an illusion, could it not be the case that both personal *and* perspectival ownership are constructed? A bundle theorist need not have any quibbles with Damasio's way of explaining the constructed nature of personal ownership. It is just that they would also add unity and unbrokenness to the mental construction of self, as indeed, Damasio himself does. There are no constraining assumptions about nirvana to rule out this possibility. Eliminating each of these rival explanations will thus be necessary in order to defend the two-tiered illusion of self, such that: (a) personal ownership is constructed/illusory and (b) perspectival ownership is not.

First, consider the specter of realism about the self. Is there reason to favor the hypothesis that the self *qua* personal owner is constructed and illusory rather than, as it purports to be, unconstructed? The advantage of the unconstructed-self hypothesis is that it does preserve the appearances. All things being equal, it is better to preserve appearances than not to: one has to tell less of a complicated story of how the appearances have the features that they do. But are all things equal here? Often what lurks behind claims that reject the veracity of appearance in favor of an illusion hypothesis are metaphysical assumptions about what items are to be allowed in one's favored ontology (consider the austerity of the Churchlands' eliminative materialism), or scientific theories about the workings of the world and its subjects (consider theories about color-illusion). In the case of the self, both factors are at play. Damasio, for instance, offers a scientifically motivated account about how the bounded aspect of self arises (and how it can disappear in pathology), which leads me to be optimistic that an entirely adequate (even if in part speculative) explanation for this aspect can be found, without recourse to a self at all.¹⁶ My metaphysical assumption is that if science can adequately explain the phenomena, using known properties and mechanisms, then it is better to avoid appealing to metaphysically extravagant alternatives to explain the data, which would multiply entities beyond what is necessary. Compared to the scientifically viable components that are needed to explain the mechanism of identification, the unconstructed self is a metaphysically extravagant entity.

¹⁶ More specifically, Damasio (1999) regards the feelings of desire-driven emotion, which fuel personal ownership and identification, to be essential in constructing the bounded sense of self. Evidence for this hypothesis is that all signs of such emotion are entirely absent during episodes of epileptic automatism and suchlike. I've argued elsewhere (Albahari 2006: ch. 8), that the close ontological relation between desire-driven emotions and the sense of self is remarkably congruent with the teachings of Buddhism (to let go emotional investment in desire-satisfaction (Pāli *taṇhā*, Sanskrit *tṛṣṇā*)).

Now to the second worry. Accepting that the *personal owner* (as a whole) is constructed, is there reason to suppose that the *perspectival owner* (a component of the personal owner) is not? I have only argued, so far, that the appearance of the perspectival owner is real—that there must seem to be a unified subject to which the stream is given. But it might turn out that the perspectival owner, even if operating as the locus of identification, is itself a mental construct, rather than unconstructed as it purports to be. In such a case it seems conceivable that its mental construction could occur through some avenue other than identification, perhaps through the innate action of memory and the imagination, as Hume claims. In such a case, the constructed synthesis of unity and unbrokenness (as it applies to witnessing perspective) could occur below the threshold of what we could ever be aware of, ruling out any hope of cutting through the illusion of self in the manner that Buddhist practice thinks possible. The advantage of such a hypothesis parallels that of the previous argument. It seems more scientifically parsimonious to favor a theory that appeals to relatively known quantities, such as thoughts, perceptions, memory, imagination, and the brain, than to quantities that are metaphysically mysterious. The presence of an intrinsically unified, invariant and unbroken witness-consciousness that qualifies the perspectival owner does, by comparison, seem to be more mysterious, not only because it resists reduction to the more familiar psychological components, but because, unlike thoughts and perceptions, it is elusive to attentive observation, making it more resistant to both scientific and introspective methods.

I've argued elsewhere (e.g. Albahari 2009) that the unified perspectival subject cannot be illusory (as conditions for the possibility of an illusion taking hold require there to be a perspectival subject). I now wish to present a different but related argument for the reality of such a (unified) subject, which focuses on conditions for the possibility of a stream of experience. Putting aside versions of eliminative materialism (that altogether deny reality to subjective life),¹⁷ advocates of a bundle theory will insist that any impression of a unified perspectival subject (should they agree to such an impression) is an illusory projection from the stream of discrete, causally connected, multi-modal experiences. Why illusory? Being a causal projection from the stream, the subject will not be unconstructed by the stream, as it purports to be. But what, then, can it mean for the stream to cause the subject (or its appearance

¹⁷ I discuss an eliminative materialist brand of no-self in Albahari (2006: 165–167).

thereof)? For A to cause B, A has to in some way exist independently of B (even if A cannot occur without B occurring at the same time). The stream of diverse experience must thus in some way exist independently of the ‘perspective’ to which it ‘appears’—a position I find untenable. Reflexive reductionists are right here to insist that such experiences, by their very nature, are subjective phenomena, first-personally given. But they stop short of admitting a real ‘first person’ (the perspectival owner) to which the experiences are given: the experiences must exist *in and of themselves* as first-personally given. Now, while I granted this possibility in the previous section, it on closer inspection borders on contradictory. For what could ‘first-personal givenness’ mean, other than ‘given to a (real) first-personal perspective’? (It cannot mean ‘given to an illusory perspective’!) ‘First-personal givenness’ is in this respect a success term: the givenness of an experience to a perspective from which it is observed entails the reality of the observational perspective to which it is given. Hence the stream of diverse experience, if first-personally given, cannot ontologically precede (and hence cause the appearance of) the perspectival owner.

But must such experiences (in order to be experiences) appear to a synchronically unified perspective? Well, if we consider any object-experience at a given time, down to the simplest sensation, it will always be possible to parse it into dimensions of which one is simultaneously aware (for example, seeing a stretch of sky entails being simultaneously aware of different locations of color). The dimensions of the experience must thus be synchronically present to the single perspective that observes it. Must the experiences appear to a diachronically unified perspective? Despite Abhidharma scholars, I cannot comprehend how an experience could present as *literally* momentary, disappearing the very moment it arises, for how then could it be said to exist at all? On this point I agree with Zahavi in his critique of Dreyfus (this volume): no matter how brief, a lived experience must have some duration, otherwise it is akin to a mathematical point, a mere abstraction. So for any stream of experience to occur, it has to occur to a point of view, and for its content to recognizably exist at all, it must be unified to that point of view for at least the specious present. In short, there can no more be a stream of experiences that ontologically precede (and give rise to) the unified minimal subject(s) to which it appears, than there can be physical objects that ontologically precede (and give rise to) the space in which they

appear. A unified minimal subject (à la perspectival owner) cannot be caused by the stream of experience: it is integral to the conscious episode itself.

Does the argument immunize longer-term diachronic unity from illusory status, where the perspectival owner seems, via witness-consciousness, to persist unbrokenly and invariantly from one specious present to the next (and hence, for the duration of each waking episode)? It does not. It is conceptually possible for there to be a stream of discrete, invariant perspectival owners, each conditioning the next, such that there appears to be only one unbroken invariant witness of the stream. Since the situation cannot be resolved by appealing directly to phenomenology, the most promising strategy may well be to defend a version of the Experience Condition, which privileges the first-person perspective. On this principle, clashes between theory and experience will be resolved in favor of the latter—provided the disputed domain of reality is restricted to the experiential. In a very interesting passage (which defends diachronic unity of—and between—the specious presents), Zahavi appeals to exactly such a principle:

Now, perhaps it could be objected that our experience of diachronic unity is after all ‘merely’ phenomenological and consequently devoid of any metaphysical impact. But to think that one can counter the phenomenological experience of unity over time with the claim that this unity is illusory and that it doesn't reveal anything about the true metaphysical nature of consciousness is to make use of the appearance-reality distinction outside its proper domain of application. This is especially so, given that the reality in question, rather than being defined in terms of some spurious mind-independence, should be understood in terms of experiential reality. For comparison, consider the case of pain. Who would deny that pain experience is sufficient for the reality of pain? To put it differently, if one wants to dispute the reality of the diachronic unity of consciousness, one should do so by means of more convincing phenomenological descriptions. To argue that the diachronic unity of consciousness is illusory because it doesn't match any unity on the subpersonal level is to misunderstand the task at hand.(Zahavi, this volume, p. 72)¹⁸

¹⁸ Zahavi would insist that the phenomenology does not support a diachronically unified *witness-consciousness*. I hope to have shown why it *is* the correct phenomenological description.

While this version of the Experience Condition needs further defense and elaboration (it should not easily allow for the existence of libertarian free will or the self, for instance), it goes some way towards establishing that moment-to-moment, unbroken, invariant witness-consciousness is real. This in turn goes some way towards establishing that the bundle theory, which bestows constructed status to such witness-consciousness, is false.

The possibility of nirvana, I argued earlier, implies a two-tiered illusion of self. In this section, I have argued that (a) the architecture of the mind is empirically compatible with a model of two-tiered self-illusion, in that it allows the absence of a sense of personal ownership to be coupled with the presence of what would seem to be perspectival ownership and (b) the best explanation for this empirical data is one that supports a two-tiered illusion of self (as opposed to self-realism or the bundle theory). This provides the possibility of nirvana with an independent measure of support.

5. Conclusion: The Challenge Ahead

Nirvana is the undoing of the self-illusion, and on the account offered, it will be a process by which the perspectival owner de-identifies with any idea held about ‘who one is’. The sense of being a bounded personal owner will be gradually eroded, freeing the (personally) ownerless consciousness intrinsic to nirvana. Having established the architecture of mind as potentially suitable for such de-identification, one might suppose that an easy task lies ahead in proving the psychological possibility of nirvana. This is far from true. A major obstacle pertains to the very case studies that served to buttress my earlier arguments. In all the cases where Damasio alludes to a sense of personal ownership being notably absent, the pathology is so severe that the patient is unable to function in the world. Ownerless consciousness has been malfunctioning consciousness. These sorts of considerations lead Damasio to conclude that ‘a state of consciousness which encompasses a sense of self as conceptualized in this book is indispensable for survival’ (Damasio 1999: 203–204).

A major challenge for those defending the psychological possibility of nirvana is thus to show how it could be possible for the sense of self to be

eroded in ways that avoid debilitating pathology. A clue may well lie in the quality of attention that is cultivated during meditation. In all the cases enlisted by Damasio, where the sense of personal ownership is entirely suspended, the quality of attention has been abnormally low (e.g. in epileptic automatism, Alzheimer's disease and akinetic mutism). The high quality of attention that is cultivated in the meditative states may thus offset the pathological side-effects, especially as Damasio notes higher-quality attention to be a reliable indicator of mental acuity (Damasio 1999; 182–183). With the mounting studies outlining the neuropsychological benefits of meditation, a measure of empirical support may well, already, be forthcoming.¹⁹

References

- Albahari, M. (2006), *Analytical Buddhism: The Two-Tiered Illusion of Self* (Houndmills: Palgrave Macmillan).
- (2009), 'Witness-Consciousness: Its Definition, Appearance and Reality', *Journal of Consciousness Studies* 16, 1: 62–84.
- Damasio, A. (1999), *The Feeling of What Happens: Body and Emotion in the Making of Consciousness* (London: William Heinemann).
- Dennett, D. C. (1991), *Consciousness Explained* (London: Penguin Books).
- Flanagan, O. (1992), *Consciousness Reconsidered* (Cambridge, MA: MIT Press).
- Gupta, B. (1998), *The Disinterested Witness: A Fragment of Advaita Vedānta Phenomenology* (Evanston, IL: Northwestern University Press).
- Hume, D. (1739–40/1978), *A Treatise of Human Nature*, ed. L. A. Selby-Bigge and P. H. Nidditch (Oxford: Oxford University Press).
- James, W. (1890/1981), *The Principles of Psychology: Vol. 1*. The Principles of Psychology (Cambridge, MA: Harvard University Press).
- Metzinger, T. (2003), *Being No One* (Cambridge, MA: MIT Press).
- Sayādaw, Mahāsi The Venerable (1994), *The Progress of Insight: A Treatise on Satipaṭṭhāna Meditation*, translated from the Pāli with notes by Nyānaponika Thera (Kandy, Sri Lanka: Buddhist Publication Society).
- Strawson, Galen (2008), 'Episodic Ethics', in *Real Materialism and Other Essays* (Oxford: Clarendon Press).

¹⁹ My thanks to Mark Siderits, Evan Thompson, Dan Zahavi and the anonymous reviewers for their critical feedback on earlier drafts. My thanks to the philosophy department at the University of Calgary for providing a venue from which to take on the critical feedback during my sabbatical leave.

- Velleman, J. D. (2002), 'Identification and Identity', in Sarah Buss (ed.), *Contours of Agency: Essays on Themes from Harry Frankfurt* (Cambridge MA: MIT Press).
- Zahavi, D. (2005a), 'Commentary on: Metzinger. T. *Being No-one: The Self-Model Theory of Subjectivity*', *Psyche* 11, 5.
- (2005b), *Subjectivity and Selfhood: Investigating the First-Person Perspective* (Cambridge, MA: MIT Press).