Philosophers' Imprint

VOLUME 14, NO. 21
JULY 2014

Insight Knowledge of No Self in Buddhism: An Epistemic Analysis

Miri Albahari

University of Western Australia

Prologue

Mary Analogue is about to give a talk on 'no self'. It is the most anticipated talk at the conference because word has it that she, like her cousin the famous colour scientist Mary, knows everything there is to know about her subject-matter.[1] At least, she knows all the theory. In particular, she knows that there is no self of a certain kind that most humans deeply buy into: a personalised and persisting centre of agency and ownership, a centre with elusive boundaries that enclose a thing of utter uniqueness and axiological salience that must be protected.[2] It is the self on behalf of which people seek to satisfy their desires, dreams and ambitions: the thing that feels emotions of pleasure (such as excitement, lust, joy) if the desires are fulfilled, and displeasure (such as anger, fear, disappointment) if they are frustrated. It is the thing that is perceived to initiate such actions to satisfy the desires.[3] Mary has closely studied a rare sector of the Buddhist community (called *arahants*) who, through years of meditation practice, are rumoured to have seen through and overcome this illusion of self. She has extracted every fact from the rumour: she knows all the intricacies of their cognitive transformation to *nibbāna* (as it's called) – how meditation changes their brain and eliminates those complex and pervasive patterns of desire-driven emotion and action that stem from an assumed identification as a solid, separate self. Amongst the

1. Despite the namesake, a complete theoretical knowledge of their subject-matter is as far as the analogy between the two Marys is supposed to go. At a later point, the two cases are contrasted. The paper will be assuming, for the sake of argument, that there is no self.
2. On the elusiveness and axiological salience of self, Gilbert Ryle writes: "He also feels, very vaguely, that whatever it is that his 'I' stands for, it is something very important and quite unique, unique in the sense that neither it, nor anything like it, belongs to anyone else." (1966, 31).
3. William James writes: "It is the home of interest—not the pleasant or the painful, not even pleasure or pain, as such, but that within us to which pleasure and pain, the pleasant and the painful, speak. It is the source of effort and attention, and the place from which appear to emanate the fiats of the will… being more incessantly there than any other single element of the mental life, the other elements end by seeming to accrete around it and to belong to it. It becomes opposed to them as the permanent is opposed to the changing and inconstant" (1890, 297–298).

emotions strikingly absent in arahants is fear: for just as our awakening from a dream disperses any fear of a dreamt-of tiger, their 'awakening' from the illusion of self disperses any fear on behalf of the formerly-assumed self-entity.

Conversely, Mary knows that a feeling of fear typically indicates a sense of the self, and that she, Mary Analogue, is afraid of public speaking. Upon mounting the podium and seeing a packed audience replete with famous philosophers, the anxiety kicks in. Struck by stage-fright she stands in the spotlight and falteringly begins to speak. This causes some people in the audience to wonder if there is not something inconsistent about Mary. There she stands, giving a paper on how there is no self – yet a sense of that very self is causing her words to tremble. People wonder: Is Mary a bit like Hume, who, after reciting philosophical arguments for no self, returns to his backgammon with bias on the imagination? Could her sense of self betray an irrational commitment to the self's existence, which contradicts (what we are supposing is) her complete theoretical knowledge that the self does not exist? Would Mary somehow improve her (already theoretically complete) knowledge that there is no self if, like those elite Buddhist practitioners, she were to eliminate the vast array of affective and behavioural dispositions that accompany the mistaken assumption that she is a self? And could such epistemic improvement cast light on what Buddhists mean when they talk about the highly esteemed event of gaining 'insight knowledge' of no self? As Mary Analogue fumbles through her talk, a member of the audience decides to write a paper addressing these epistemological questions with a view to analysing a topic that has been given little attention in Western analytic philosophy. Could there be anything epistemically distinctive, and indeed profound, about the gaining of so-called *insight knowledge* into the reality of no self? And could anything about such analysis illuminate the epistemic structure of a wider range of cases, such as the overcoming of a phobia one knows is irrational?

0. Introduction

Nibbāna is the *summon bonum* of Buddhist practice.[4] Putting aside mystical descriptions, a central tenet of *nibbāna*, as described in early Buddhist tradition, is that it yields a complete understanding of the truth of no-self. Sometimes called 'awakening', the realisation is said to permanently free the aspirant from the affective, behavioural, and motivational drives that stem from having a sense of self. It is said to be a state of utmost contentment and equanimity, with no capacity to mentally suffer (hence no unpleasant emotions such as fear or gloom). Attention is sharp and never lost in thought. There is immense compassion towards people's suffering, yet without any attachment to outcomes. There is no identification with elements of the mind and body, which would give rise to such thoughts as 'this is *me*, this is *mine*, this is *my* action'.[5] The process of understanding the reality of selflessness is thus

4. My interpretation of Buddhism draws primarily upon early Buddhist teachings from the Pāli *suttas* (discourses between the historical Buddha and his disciples) and as expounded in the work of leading Buddhist scholars such as Bhikkhu Bodhi. I thus use Pāli spelling in all the Buddhist terminology.

5. Elsewhere I introduce a distinction between what I call 'perspectival' and 'personal' ownership (Albahari 2006, 2011). Perspectival ownership is the sort of mine-ness neutrally borne towards objects that happen to appear uniquely to one's perspective (such as thoughts, feelings, perceptions and bodily actions); personal ownership is an emotionally invested mine-ness that is reciprocally borne from identifying, amongst other things, with perspectivally owned elements of one's body-mind as 'me' and 'who I am'. In the Pāli suttas, the Buddha alludes to identification (evidenced by thoughts of 'this is me') along with reciprocal feelings of personal ownership ('this is mine') as being central to the sense of self. Thus one encounters such passages as "'Bhikkhus, there being a self, would there be for me what belongs to a self?' –'Yes, venerable sir.' – 'Or, there being what belongs to a self, would there be for me a self?' –'Yes venerable sir'" – (MN 22, 1995, transl. Ñāṇamoli and Bodhi). One should thus aspire to a state where no longer identifies with or feels personal ownership towards any facet of their psychophysical existence, hence: "What is nonself should be seen as it really is with correct wisdom thus: 'This is not mine, this I am not, this is not my self'." (MN 22 and SN 35, 1995, transl. Bodhi). The following passage starkly illustrates the complete lack of emotionally invested identification and personal ownership that should eventually be harboured towards each element of psycho-physical existence:

> "…'Bhikkhus, what do you think? If people carried off the grass, sticks, branches, and leaves in this Jeta Grove, or burned them, would you think: 'People are carrying off or burning us or doing what they like

said, in the Buddhist tradition, to be not merely intellectual, but deeply transformative — integrally connected to the experience of eliminating the sense of self and its psychological structures. Such structures are said to sustain mental 'defilements', such as preferences and aversions, which make one attached to things being one way rather than another, causing us to suffer when our desires are frustrated. On the insight into selflessness, a leading scholar monk, Bhikkhu Bodhi, writes:

> Of these cognitive distortions, the most deeply grounded and resistant is the delusion of self, the idea that at the core of our being there exists a truly established "I" with which we are essentially identified. This notion of self, the Buddha teaches, is an error, a mere presupposition lacking a real referent. Yet, though a mere presupposition, the idea of self is not inconsequential… Because we make the view of self the lookout point from which we survey the world, our minds divide everything up into the dualities of "I" and "not I," what is "mine" and what is "not mine." Then, trapped in these dichotomies, we fall victim to the defilements they breed, the urges to grasp and destroy, and finally to the suffering that inevitably follows… To free ourselves from all defilements and suffering, the illusion of selfhood that sustains them has to be dispelled, exploded by the realization of selflessness. Precisely this is the task set for the development of wisdom… wisdom removes the veils of distortion, enabling us to see phenomena in their fundamental mode of being with the vivacity of direct perception. The training in wisdom centers on the development of insight *(vipassanā-bhavana)*, a deep and comprehensive seeing into the nature of existence which fathoms the truth of our being in the only sphere where it is directly accessible to us, namely, in our own experience. [1994, 56]

with us?' – 'No, venerable sir. Why not? Because that is neither our self nor what belongs to our self.' – '…What is it that is not yours? Material form is not yours…Feeling is not yours…Perception is not yours… Formations are not yours…Consciousness is not yours…Abandon it. When you have abandoned it, that will lead to your welfare and happiness for a long time.'" (MN 22). For more on how identification could contribute to the sense of self, see note 16.

While the general topic of insight knowledge in Buddhism has received little attention in Western philosophy, Galen Strawson in his re-released book *Freedom and Belief* writes:

> It is not implausible to suppose that Buddhist monks and mystics have succeeded in altering quite profoundly their experience of themselves as acting, thinking, and feeling beings. Nor…is it implausible to say that they have in so doing achieved what is in certain respects a more correct view of the world…[2010, 103] [6]

This paper offers an analysis of what it could mean, in epistemic terms, to arrive at a "more correct view of the world" through the profound alteration of lived experience – or of what is termed, by Buddhists, as 'insight' (*vipassanā*), or more broadly, 'wisdom' (*paññā*).[7] While Buddhist traditions allude to varying targets and degrees of insight knowledge, my analysis will focus on what is widely agreed, in early Buddhism, to be an insight of the most profound in nature: that of fully apprehending the reality of there being no self, through the attainment of *nibbāna*. My analysis need not assume that the full purported insight into no self must occur *all at once* from a stage in which the practitioner has a

6. Owen Flanagan has in a recent book described Buddhist wisdom as "absorb[ing] and internaliz[ing] a certain metaphysic of self" (2011, 131). His treatment of the topic, however, focuses on the psychological and ethical dimension — how wisdom may diminish suffering/desire and promote happiness/compassion — rather than on an epistemic analysis.

7. While these two terms indicate differences in emphasis (*vipassanā* is more associated with the activity of cutting through delusion, and *paññā* with the resulting wisdom), I will, for purposes of this paper, use the English term 'insight' or 'insight knowledge' to cover both these aspects. I will use the term 'awakening' to refer to the event of having attained *nibbāna*, and *arahant* (sometime un-italicized) to refer to one who has awakened.

definitive sense of the self. If the attainment of *nibbāna* occurs over a series of smaller stages or breakthroughs, my inquiry will compare the stages of where the practitioner definitively has a sense of the self with the final stage at which all traces of the self-illusion have vanished.

My question is thus: assuming that there is no self and that it is possible to lose the sense of self in a way that retains normal psychological functions, how might we articulate and explain the appearance of epistemic progress that occurs when the practitioner is said to gain full insight into the reality of no self? Buddhist tradition puts much emphasis on the claim that insight knowledge is not (or not merely) theoretical knowledge, but is knowledge of a kind that is gained via experience. When approaching the issue, it is thus instructive to compare Mary Analogue with the subject of Frank Jackson's (1986) famous thought-experiment, 'Mary the Colour Scientist'. Raised from birth in a black-and-white room, Mary acquires complete physical knowledge about the physics and physiology of colour and colour vision. Upon release from her cell, she sees colour for the first time, prompting in philosophers (amongst other conclusions) the widespread intuition that she makes epistemic progress that goes beyond her theoretical knowledge. Many will claim that Mary now knows, in some experiential or practical sense, *what it is like* to see colour. If Mary Analogue were to be liberated from the illusion of self, for the first time experiencing a mind completely freed from the illusion, would any epistemic progress be best described along the *same lines* as Mary the colour scientist (such that she now knows *what it is like* to be freed from the illusion of self), or would there be also something else that is distinctive about her epistemic improvement?

While I surmise that there would indeed be parallels with the epistemic progress of Mary the colour scientist (if such there be), I will propose that Mary Analogue's impression of having a *more correct view of the world* — typified in reports from Buddhist traditions — would indicate that there is something quite distinctive, as well, about her progress. I offer an analysis of what, at least in part, this distinctive kind of epistemic improvement could amount to. The core of my analysis will expand upon the Strawsonian/Buddhist description of insight as 'overcoming a delusion'. Delusions, at least the non-clinical kind, are commonly thought to be types of stubborn, false belief. Building on this idea, I hypothesise that the gaining of insight knowledge, through losing the sense of self (of a particular nature), would involve the uprooting of a deep-seated and reflexive *false belief* that one is a self, along with the re-alignment and integration of one's emotional, cognitive and behavioural dispositions in accordance with the correct belief that there is no such self. If the correct belief is already a component of the subject's existing (theoretically-based) propositional knowledge — as I am assuming is the case with Mary Analogue and most Buddhist practitioners — there will be the loss of a false 'action-based' belief that *contradicts* the doxastic component of this knowledge. Through subsequent doxastic integration of action-based with 'reflective' belief, I contend that the subject's propositional knowledge, that there is no self, will have greatly improved in quality.[8] This type of experientially based epistemic progress is to be distinguished from that had by Mary the colour scientist; it is more than simply coming to *know what it is like* to experience a mind freed from the illusion of self.

The success of such a proposal will rely on the truth of at least three claims, namely that: (1) the sense of self is doxastically anchored in an action-based belief, (2) if doxastic, one can simultaneously harbour such an action-based belief (that one is a self) with an opposing reflective belief (a component of the propositional knowledge that that there is no such self), and (3) replacing the false action-based belief with one that doxastically integrates with the correct reflective belief would improve the quality of one's existing propositional knowledge that there is no self. While there will not be room to fully defend each of these claims, I will offer some lines of argument that can be pursued in their defence, with the suggestion that upsurping the sense

8. The terms 'action-based belief' and 'reflective belief' will be properly explained in section 3; for now, it is enough to note their respective connection with patterns of action/emotion versus reflective endorsement.

of self in this way involves the alteration of a deeply foundational 'framework belief'.

The general analysis of knowledge-improvement, if correct, will not be confined to the gaining of Buddhist insight knowledge, but should apply to other cases of where there is dissolution of conflict between (the doxastic component of) propositional knowledge and recalcitrant beliefs. I will suggest that in suitably doxastic instances of (say) losing a phobia, superstition or clinical delusion, the subject's propositional knowledge (*e.g.* that feathers are not dangerous) will similarly have improved through the replacement of a contradicting, false action-based belief (*e.g.* that feathers are dangerous) with a correct belief that integrates with the doxastic component of the subject's existing propositional knowledge.

While extending the analysis of knowledge-improvement to a wider range of cases may be viewed as an advantage of the account, it also raises the question of whether there is anything really distinctive, after all, about the gaining of so-called insight knowledge. Could there be something substantive behind the fact that the nomenclature 'insight' (or 'wisdom') occurs within Buddhist traditions, rather than in connection with the loss of the phobias or suchlike? Indeed there would appear to be. The 'insight' arising from overcoming the delusion of self is said to carry an aura of profundity; it is a cognitive transformation that deeply, globally and irrevocably shifts one's entire perspective on the world – befitting the alteration of a fundamental framework belief. I will offer some empirically based speculations on what could account for this profound cognitive shift, insofar as it explains why the level of insight into the reality of no self may be distinguished, at least in degree, from that of the more mundane cases.

Doing full justice to the topic of Buddhist insight knowledge will be a far lengthier enterprise than what can be covered within the scope of a single paper. What I hope to provide is some philosophical traction to the idea of gaining insight knowledge of no-self: namely, a core proposal of its epistemic structure, along with an indication of the direction that further research may take in defending key claims on which the proposal rests. I do not purport to provide a complete overview on the nature of insight knowledge, nor do I pretend to even touch upon everything that could be profound about it. Moreover, offering such a hypothesis on the gaining of insight knowledge into no-self will require making, for the sake of argument, several provisional and contentious assumptions. Of these, none are so contentious as the assumption that *nibbāna — qua* losing the sense of self (and associated affective and behavioural drives) while retaining or enhancing mental acuity and well-being — is psychologically possible. Despite Strawson's optimism that such a supposition is "not implausible", something must be said to allay the legitimate concern that it is so implausible as to demotivate the project from the outset — at least as part of a serious inquiry into what human cognition is capable of (as opposed to a mere exercise in speculative logic, akin to analysing how many angels can fit on the head of a pin).

Section 1 will thus be devoted to expounding upon and addressing this major concern, before turning, more briefly, to relatively less contentious presuppositions that are assumed for purposes of my discussion: that there is no self, and that we have a sense of the self. While addressing these presuppositions will not, of course, justify them, my goal is to show that they are not so obviously implausible as to arrest the project before it can get started. I aim to show that philosophical inquiry into gaining insight knowledge of no-self, as described in early Buddhist teaching, is of genuine relevance to the contemporary fields of knowledge, mind and cognition. It is hoped that this exercise will also make more concrete the overall context of inquiry, such as how it sits with current empirical research, as well as elucidate the difference between the central notions of *self* and *sense of self* as they stand in relation to the non-existence of self. Unless spelt out in sufficient detail, subsequent discussion about the epistemic benefits of 'losing the sense of self' will have little to hang on.

The remainder of the paper has already been foreshadowed. In section 2, I introduce, in more detail, the comparison of Mary Analogue with Mary the Colour Scientist, via the question: "what distinctive

epistemic dimension could the gaining of so-called *insight-knowledge* add to already perfect theoretical knowledge of the proposition 'there is no self'"? In developing this component of the account, I introduce a further passage by Strawson from which I draw out my specific proposal. Then, in section 3, I offer preliminary arguments for the three further claims upon which this proposal depends, which would allow for knowledge-improvement through doxastic integration. In section 4, I describe how the account of knowledge-improvement could extend to other cases, including the loss of phobias, clinical delusions, and superstitions. In section 5, I offer some empirical speculations, based upon the account of self in section 1, on what could make Mary Analogue's knowledge-improvement — as opposed to (say) the loss of a phobia — profoundly insightful. I conclude by briefly considering how the account might bear upon cases where a subject lacks initial knowledge that there is no self.

1. The empirical viability of the nibbānic hypothesis

Let us call the proposal that *losing the sense of self whilst retaining or enhancing mental acuity is psychologically possible* the 'nibbānic hypothesis'. A central concern for the nibbānic hypothesis lies in evidence from scientific quarters to suggest that losing the sense of being a separate, axiologically salient self, along with attendant self-regarding emotional and behavioural patterns — even if the self is an illusion — is sure to result in sub-human rather than super-human states. One of the most carefully worked out and influential hypotheses about the origins and neurological underpinnings of the self-sense is to be found in the work of neurologist Antonio Damasio (1999, 2012).[9] From his studies of a number of such cases where the sense of self has been suspended, involving pathologies such as akinetic mutism and epileptic automatism, Damasio concludes:

> When the mental aspect of self is suspended, the advantages of consciousness soon disappear. Individual life regulation is no longer possible in a complex environment. …In fact, left to their own devices, death would ensure in a matter of hours because bodily maintenance would collapse. This, and comparable examples, would suggest that a state of consciousness which encompasses a sense of self as conceptualised in this book is indispensable for survival. [1999, 304–305][10]

How, in more detail, is this self to be defined, and how is it to be distinguished from the sense of self? How might we understand the claim that the sense of self is the sort of thing that exists, while the self is not? And how might such a sense of self be seen, on Damasio's view, as essential to survival?

While there are many notions of self in the literature, the core notion of self at stake in Damasio's work, in Buddhism, and in that of several Western philosophers is that of "you as observer or knower of the things observed, […] you as owner of thoughts formed in your perspective, you as potential agent on the scene" (1999, 127).[11] This owner/observer/agent is a personalised and persisting entity: a unique, uni-

9. Similar ideas have been proposed by neurologist Jaak Panksepp (1998).

10. With regard to the necessity of the mental construction of self, Daniel Dennett also writes: "Stripped of it, an individual human being is as incomplete as a bird without its features, a turtle without its shell."(1991, 416). See also note 18 on Panksepp (1998).

11. Western philosophers who have also denied the existence of a self of this description (or something close) include Hume (1739/1978), James (1890/1981), Parfit (1984), Dennett (1991), Flanagan (1992), Metzinger (2003), and the author (2006). Elsewhere I provide detailed evidence and argument for the view that Damasio and some of these Western philosophers are dealing the same or very similar notion of self that can be gleaned from suttas in the Buddhist Pāli Canon (Albahari, 2006). I also offer a comparative analysis of how, according to Buddhist and Western traditions, the illusion of such a self may be said to arise. From this, it becomes apparent that not everyone, even within a particular tradition, agrees upon how the sense of self gets constructed. Dennett, for instance, has a more linguistic emphasis than Damasio, seeing the self as a "centre of narrative gravity". In a later book Damasio writes: "There is indeed a self, but it is a process, not a thing, and the process is present at all times when we are presumed to be conscious" (2012, 8). This seems mainly to be a semantic shift: what he is describing is the neurological (etc.) process underpinning the *sense* of self, a process he also ascribes to pre-linguistic

fied and bounded locus of agency that underlies and is somehow generative of our thoughts and experiences. This self is described as systematically elusive to its own observation. While able to turn its focal awareness onto its thoughts and experiences, as well as aspects of the wider world, the supposed self can never seem to directly observe itself in this manner. The elusiveness is what essentially distinguishes the self as a type of *subject* rather than object in the world, aligning it with the observer rather than with the things directly observed. Yet the self still seems reflexively and peripherally aware of its own presence, such that an individual's experience is not confined to objects of awareness, but seems divided into the observing subject (*qua* self) and the observed objects. This subtle, reflexive feeling that we have of our own presence *as* such a bounded thought-antecedent self, as something distinct and separate from its surrounding environment, is what is referred to as the 'sense of self'.[12]

creatures. It remains the case that *qua* bounded observer/owner/actor, his theory implies there is no such (unconstructed) thing as a self.

12. It is a presupposition of this paper that we do indeed have a sense of being an entity with the above-listed features. Still, something should be said about how, given that we cannot directly introspect and 'read off' the characteristics of our supposed selves, we can arrive at the list of features that we supposedly ascribe to ourselves. Much of the content has to be inferred indirectly, through reflecting on our modes of interaction with the world, including our likely motivations, emotions and behaviours. Take Mary Analogue's rising fear at the upcoming talk. This indicates, arguably, that she deeply identifies as the person who is about to give the speech, such that she assumes it to be the numerically same being as the one now undergoing the anxiety, implying an assumption of personalised, uninterrupted *persistence* over time. Suppose that she berates herself for being abrupt with a questioner in the audience, thinking 'I should not have said that!' This guilt would indicate not only an assumed persistence, but the fact that she thinks it possible, all else being equal, that she *could* have acted otherwise. This arguably implies that she takes herself to be an *agent* with libertarian free-will. In the current section I elaborate on how the assumption of boundedness can be inferred from (and indeed constructed by) cognitions that amplify an ongoing reflexive concern for one's own welfare, although I argue elsewhere that boundedness can be inferred from a multitude of factors, including the sense of agency. For a detailed account and defence of how the entire list of features gets ascribed to the self, see Albahari 2006, 2011.

Now, those who deny the existence of such a self do not usually deny that the sense of self is real, any more than denying that the two lines in a Muller-Lyer Illusion are of uneven length involves denying the appearance of such lines. What is held to lack reality — at least in its entirety — is rather the very thing that we have a reflexive sense of being: a self, with all the enlisted features, that is wholly antecedent to and unconstructed by the thoughts and experiences that it appears to own or generate. The idea is that instead of being anchored in a thought-antecedent self, as they subjectively appear to be, at least *some* features of the thing that we reflexively and unwittingly take ourselves to be (via the sense of self) — such features as boundedness, persistence, agency, unity, axiological salience — turn out to be wholly or partially generated by the very thoughts and experiences that the self seems to own or generate. Put simply, the self (with those features) does not, as it purports to, think the thoughts; instead, the thoughts think (those features of) the self. The mismatch between appearance and reality is what makes the self an illusion.

Denying the existence of self does not entail, then, denying the reality of *every feature* ascribed to the self, a consequence that some would find implausible. Non-illusory, unconstructed features ascribed to the self can survive dissolution of the self-illusion, hence the locution 'losing the sense of self' should be read as 'losing the sense of those illusory features ascribed to the self'.[13] Now, despite disagreement over the range of features said to be mentally constructed, both Damasio and the scholars of early Buddhism are likely to converge on at least the following. They will agree that the *uniquely personalised boundary* that separates self from the rest of the world (which I call 'boundedness') — the feature that makes 'me' seem like a distinctly separate, unique, axiologically salient thing in relation to the world — is mentally constructed and hence, illusory (because it

13. For instance, I argue elsewhere that the feature of conscious, unified awareness (*modus operandi* of 'observer') cannot be mentally constructed and hence illusory (Albahari 2006, 2011).

purports to *not* be mentally constructed).[14] This personalised boundary between self and the world is much psychologically 'thicker' than that which comes from merely occupying an embodied, first-person perspective on the world.[15] As Gilbert Ryle puts it, "He also feels, very vaguely, that whatever it is that his 'I' stands for, it is something very important and quite unique, unique in the sense that neither it, nor anything like it, belongs to anyone else" (1966, 31). The feeling of importance attached to being 'this very thing' may be summoned by the reaction of horror around the prospect of being replaced by a psychophysical replica who will go on to live your life (an idea explored in Parfit, 1984, 199–201).[16]

14. While I regard such boundedness as central to the self-illusion in Buddhism, other authors may take other features of the self to be primary targets for Buddhist meditative practice. Galen Strawson holds "regarding [oneself] and others as truly self-determining sources of action" (2010, 103) – viz., an agent of libertarian freewill – to be a such a target. While I remain neutral on this point, if Strawson is correct, my overall philosophical analysis (of what it is to gain insight into no self) will still apply.

15. The passive boundary that comes from merely occupying a first-person embodied perspective goes reciprocally with what I refer to in note 5 as 'perspectival ownership'. 'Boundedness' pertains to the emotionally invested, more actively (albeit elusively) sensed boundary that arises with identification (more in note 16), reciprocal with what I've called 'personal ownership'. It is alluded to in the suttas in note 5 and in Bhikkhu Bodhi: "Because we make the view of self the lookout point from which we survey the world, our minds divide everything up into the dualities of "I" and "not I," what is "mine" and what is "not mine."" (1994, 56). I surmise that Damasio's patients who have lost their sense of the bounded self through pathology would still have a passive first-person, embodied perspective on the world; they'd lack any sense of bounded identity in relation *to* this perspective. So, for that matter, would Buddhist arahants – although the level of their cognition would vastly differ from the pathological cases.

16. In more detail, how might identification contribute to (and provide evidence for) the sense of boundedness and axiological salience? Identification occurs as one appropriates various ideas (*e.g.* those pertaining to specific attributes such as body, gender, age, race, character traits, preferences, profession, or to common modes of interaction such as observer, owner or agent) to one's perspective, so that the world is approached and thought about through their assumed, reflexive filter. As J. David Velleman puts it: "If there is a part of your personality with which you necessarily think about things, then it will be your mental standpoint, always presenting a reflexive aspect to your thought" (2002, 114). The numerous mantles of identification help to reify what I think is a non-illusory locus of perspectival consciousness into a personalised, solid-seeming *me*, thereby sharpening and exaggerating the sense of boundary between one's assumed existence and the world. The emotional investment in these mantles of identity would, moreover, reinforce the feeling of boundedness by feeding into an ongoing asymmetrical concern about one's welfare (alluded to by Damasio), thereby presupposing a distinct and important thing on behalf of which one is concerned. The process of awakening can in view of this also be understood as one which dismantles the layers of identity and accompanying self-concern; I say more about this soon, and in Part 5.

17. My presentation of Damasio's position here is simplified, focusing mainly on his account of how emotions and feelings help to construct the feature of boundedness, and ignoring other technicalities about the construction of self, including his account of how the organism's relation to the environment is mapped in the brain. Damasio thinks that all features of self are constructed, not just boundedness.

According to Damasio, this subtle sense of ourselves as something psychologically bounded, separate, unique, and important is both mentally constructed and critical to the survival of our organism's biological boundaries. To see how this may be so, one must consider, at least in outline, the central role that Damasio ascribes to emotion in constructing the bounded self.[17] Damasio uses the term 'emotion' to refer specifically to the body's complex set of stereotyped chemical/neural responses to the environment; 'feelings' refers to the familiar subjective side of these emotions. Through conditioning, emotions become associated both with objects in the environment, and with patterns of motor response such that:

> Memories of an object that was once actually perceived include not only records of the sensory aspects of the object…but also records of the motor adjustments that necessarily accompanied the gathering of the sensory signals… and the obligate emotional reaction to the object. As a consequence, when we recall an object…we retrieve not just sensory data but also accompanying motor and emotional data…the past reactions of the organism to that object. [1999, 161]

While it is evident that various objects and situations can regularly induce strongly felt emotional reactions — to which he gives the familiar names 'fear', 'anger', 'hope', etc.' — recent studies have substantiated Damasio's contention that even relatively neutral objects such as tables and coffee cups produce measurable 'micro-valences' (Lebrecht et al, 2012) . These subtle affective feelings of pleasantness or unpleasantness may lie below the threshold of conscious awareness, but they nevertheless prime perception, helping us to more quickly identify and act upon sources of perceived harm and benefit. Damasio holds, then, that the function of emotion — both felt and unconscious — is to reliably direct attention to the environment in ways that ready us to act so as to avoid harm and procure advantage. On the connection between attention and emotion, he writes:

> Emotion is critical for the appropriate direction of attention since it provides an automated signal about the organism's past experience with given objects and thus provides a basis for assigning or withholding attention relative to a given object. [1999, 273]

Damasio maintains that the felt, attention-directing emotions become far more motivating if the organism psychologically identifies as a bounded, axiologically salient self *on behalf* of which the emotional reactions are felt.[18] The wordless impression of being such a self is synchronically generated, he claims, by the ongoing stream of thoughts, images and felt emotions (attended and unattended), whose content represents, as part of its narrative, *not only the objects perceived and acted upon but also the organism perceiving and acting upon them* (1999, 93, 188–192). The biological organism is in other words represented *via* the subjective elusive impression of a background agential self: a *protagonist* who engages with the world in such a way that it is driven by "an individual concern which permeates all aspects of thought-processing, focuses all problem-solving activities, and inspires the ensuing solution" (1999, 304). Identifying as such a protagonist makes the organism reflexively *care* about welfare, by fueling an ongoing sense of urgency in the service of helping the organism to automatically think and act in such a way as to preserve its biological boundaries. The feeling of being a thing that must be protected greatly exaggerates the (assumed) boundary between minded organism and environment.[19]

In relation to Buddhist thought, Davis and Thompson have noted that the notion of valence (the purely affective component of what Damasio calls 'feeling') has a direct analogue in the Buddhist notion of *vedanā*: the raw feeling of pleasure, displeasure or neutrality as manifested through the six sense modalities (including the mind). They write:

> In the case of both concepts, valence and *vedanā*, the feeling tone of pleasant versus unpleasant is closely related

18. On the close connection between our basic, elusive sense of bounded agential self and sensory/affective/motor responses, there appears to be some concordance between Damasio and Panksepp. Like Damasio, Panksepp regards the core sense of self, viz., "our ego, the feeling of 'will' or 'I-ness'" to be rooted in deep, evolutionarily primitive structures of the brain that serve as the first point of contact for the intermixing of "motor maps (*i.e.*, body schema), sensory maps (world schema) and emotional maps (value schema)" (1998, 300). The interaction of these structures — involving circuits that likely "first represented the body as an intrinsic and coherent whole" and through which "a variety of sensory stimuli become hedonically valenced" — feeds into "*that ineffable feeling of experiencing oneself as an active agent in the perceived events of the world*" (1998, 310, his italics). And like Damasio, Panksepp holds that a breakdown of the primitive neural circuits that subtend the basic sense of self (with their connection to affective and motor responses) would probably involve "major deficits" in higher cerebral functions (1998, 314).

19. On Damasio's theory (and indeed Panksepp's), the more complex the environmental pressures, the more developed the sense of self will have to be to cope with them. Most animals and humans have the basic sense of self (or "core self") so far discussed: an unreflective sense of identity as an agent/owner/observer that is bounded and separate from its environment, able to cognise its immediate future and past. Once longer-term planning and decision-making become advantageous, requiring a conscious representation of oneself as the subject of remembered and imagined outcomes, the sense of boundedness and identity over time gets greatly enhanced into what Damasio calls the "autobiographical self", with many additional layers of identification such as those mentioned in note 16.

> to action tendencies of approach versus avoidance. From the modern neuroscience perspective, the bodily responses constitutive of an emotion, including an emotion's valence and action tendency, can be activated even when we do not report consciously feeling the emotion… Moreover, recent work has shown that such implicit affect valence is not limited to emotional episodes and influences decision-making on everything from consumer choices to moral judgement…This understanding of the pervasive role of affect valence in human psychology finds a parallel in the Buddhist suggestion that *vedanā* is present with every mental state, not just those Western psychology includes under the emotions. [2013, 587–588]

It is important, nonetheless, not to conflate *vedanā* — raw sensory pleasant, neutral or unpleasant hedonic tone — with the action tendencies of approach and avoidance towards those hedonic tones (feeling an emotion would usually involve a mixture of these). According to early Buddhist teaching, the quality of *vedanā* helps condition our mental reactions of preference or aversion (*taṇhā*) to such sensory stimuli, which influences (and is influenced by) our perception of things (*saññā*) as desirable or undesirable, as well as the arising of mental formations (*saṅkhāra*) that include volitional tendencies (*cetanā*) and object-specific grasping (*upādāna*). From these *taṇhā*-driven mental formations arise thoughts and feelings of 'me' and 'mine': manifestations of what I've described (in notes 5, 15, and 16) as the emotionally invested personal ownership and identification, central to the early Buddhist notion of self. The root cause of such I-thoughts lies in both the sense of self that they perpetuate, and in *taṇhā*. *Taṇhā* corresponds closely to Damasio's notion of the permeating individual concern. Often translated as craving or attachment, it is the disposition to emotionally invest in the satisfaction of desire. It is the underlying current of desire to prefer that the world be one way rather than another, such that one is made happier when the desire is fulfilled (in reaction to pleasant *vedanā*) — and less happy when it is not (unpleasant *vedanā*). *Taṇhā* is behind the constant drive, conscious and unconscious, to bring states of affairs into line with one's preferences. While the resulting thoughts and emotions of satisfaction or frustration seem to be experienced on behalf of a personal self, and are perpetuated so long as one has the sense of being such a self, Buddhism contends, like Damasio, that there is no actual such self: no thought-antecedent, thought-generating, axiologically salient corner of the world — an underlying, separate 'I' that will stand to benefit or lose from the situation at hand. The sense of boundedness that seems to separate the self as a salient thing from the rest of the world (including one's thoughts) is held to immediately stem, instead, from the very stream of *taṇhā*-driven thought and emotion that is assumed to originate in the self.[20]

Yet despite convergence over how the boundedness of self is psychologically constructed, Buddhist thought radically diverges from Damasio (and Panksepp — see note 18) over the *necessity* of the sense of this self for autonomous human existence. It is not that Buddhist tradition would reject Damasio's theory about the origin of the self-illusion; it may well accept that the sense of bounded self, with its accompanying desire-driven emotions, evolved as a complex survival mechanism that continues to serve its important biological function. But as was evident in the passage from Bhikkhu Bodhi, Buddhism does not regard a well-functioning human mind as *having* to be animated by this desire-driven sense of self. The core teaching expressed in what is known as the 'Noble Eightfold Path' (the 'Fourth Noble Truth') along with the doctrine of Dependent Origination, maintains that it is possible, via meditative practice, to break down the conditioning between

20. Note: the idea that the sense of self (perhaps also a belief that one is a self) perpetuates I-thoughts should not be confused with the mistaken idea that it is the actual thought-antecedent *self* that perpetuates those thoughts. Compare: Jim is cowering under the bed because he senses that aliens are watching him. It is entirely the sense (perhaps also a belief) that aliens are watching him that makes Jim cower, rather than anything about an actual situation of aliens watching him — there is no such situation.

vedanā and *taṇhā*.[21] As *taṇhā* is weakened, one reacts with decreasing preference or aversion to pleasant and unpleasant *vedanā*. This in turn weakens the illusion of self (with thoughts of 'I' and 'mine') that both depends upon and helps generate these emotionally invested reactions.[22] Such practice is said to culminate in the exulted *nibbāna*: the final flash of insight that burns out *taṇhā* and the sense of self for good. By contrast, Damasio and Panksepp regard the emotionally-invested sense of self to be universally critical to survival. The contention is more than just theoretical; we have seen that Damasio appeals to a number of clinical studies: subjects who, through various neuropathologies, have lost the sense of bounded self (along with all manifestations of emotion) and are unable to fend for themselves.

How serious are these concerns for the nibbānic hypothesis? It is worth remembering that Damasio's theory, while empirically supported in many of its details, is a *hypothesis* rather than verified fact: and the cases where he documents a suspension of the self-sense all occur in the context of trauma to the brain, such as stroke. The inability to act autonomously may thus be as much to do with the trauma as with the loss of the sense of self. More significantly, Damasio has not studied cases of where subjects have purportedly lost or diminished the sense of self via the sort of meditative practices detailed in Buddhist traditions. Perhaps the *way* in which the sense of bounded self is eroded is crucial to harnessing or developing capacities of the mind and brain that may permit autonomous survival *sans* the sense of bounded self.

As it happens, Buddhist meditative practices and the cognitive and neural correlates of having practiced for anything from a few minutes to over 44,000 hours have increasingly been studied in laboratory conditions. While still at an early stage, some of the findings to emerge are promising. For example, studies have indicated that increases in hours of meditation positively correlates with decreases of activity in neural correlates associated with self-narrative (mid-cortical structures)[23] and with fear, depression, and other self-concerning emotions (*e.g.*, the amygdala).[24] There is also mounting neuropsychological evidence to suggest that, for advanced practitioners of meditation, there is a substantial increase in the level, quality, and ease of attention that is paid to ongoing stimuli (rather than only stimuli that is of selective interest to the self).[25] These findings are significant for at least two reasons.

21. The Four Noble Truths, foundational to Buddhism, are part of the first discourse of the Buddha (SN 56.11, 1995, transl. Bodhi) and can be summarised as: (1) Suffering (*dukkha*) exists, (2) The origin of *dukkha* is *taṇhā*, (3) The cessation of *dukkha* lies in the cessation of *taṇhā*, and (4) There is a path to the cessation of *dukkha* (and to *nibbāna*): the Noble Eightfold Path, involving the practices of insight-wisdom (*paññā*), meditation (*samādhi*), and virtue (*sīla*). The doctrine of Dependent Origination (*Paṭiccasamuppāda*, SN 12) identifies 12 proximal links in the cycle of birth and death (*saṃsāra*), expressing the central Buddhist idea that everything which arises depends on multiple conditions. While many links are seen as passively determined (*e.g.*, "with six sense bases (saḷāyatana) as condition, contact (*phassa*) comes to be; with contact as condition, feeling (*vedanā*) comes to be"), the link "with *vedanā* as condition, *taṇhā* comes to be" is recognised as one that can be actively broken.

22. Hence, undermining the sense of self—based on a form of ignorance, *avijjā*—in turn diminishes *taṇhā*, both of which are purportedly at the root of mental suffering for the living practitioner. As mental suffering is finally eliminated through insight, unpleasant *vedanā* will be confined to only physical (not mental) suffering. And as the arahant is said to not be reborn into the cycle of birth and death (*saṃsāra*), physical suffering eventually ceases.

23. "While freely engaging in self-referential thought, individuals [in control groups] exhibited distinct engagement of cortical midline structures…and posterior cingulate cortices, regions associated with the affective appraisal of events as good or bad for the self. …By contrast, mindfulness practitioners engaging in a metacognitive process exhibited a pronounced shift away from midline cortical activation [and toward] sensory representations in the insula and secondary somatosensory cortices. These regions may support more detached, objective interoceptive, and somatic awareness that may serve as the primitive sensory representations of the 'self'" (Farb, Anderson et al, 2010, 26).

24. In a paper summarising a number of studies, Davidson and Lutz write: "Expert meditators also showed less activation than novices in the amygdala during FA [focused attention] meditation in response to emotional sounds. Activation in this affective region correlated negatively with hours of practice in life…This finding may support the idea that advanced levels of concentration are associated with a significant decrease in emotionally reactive behaviours that are incompatible with stability of concentration" (2008, 173).

25. "A recent study used fMRI to interrogate the neural correlates of FA [focused attention] meditation in experts and novices.…Whereas expert meditators with an average of 19,000 practice hours showed stronger activation in these areas than the novices, expert meditators with an average of 44,000 practice

First, the movement from increased to decreased emotional activity, and from a selective towards an impartial pattern of attention, serves as evidence that the sense of self —correlative with selective *taṇhā*-driven attention being paid to objects of personal significance — can be eroded. Second, in those contrasting pathological cases where the sense of self is severely compromised, the level of attention is usually abnormally low, something that Damasio views as indicative of pathology. Perhaps there comes a stage where a sustained, high level of effortless, unbiased attention to one's surroundings helps circumvent any pathology to the degree that it eventually *substitutes* the visceral feeling of self-concern that, with our usual (lower) levels of attention, is needed to keep the organism out of danger. Of course these reflections on the possibility of altogether eliminating the sense of self via meditative practice are speculative, but rather than being based merely upon religious conviction and untestable, they extrapolate from scientific studies and are testable.[26] The nibbānic hypothesis, while still requiring further evidence, is not an unreasonable one.

The other immediate provisional assumptions — that there is no self and that we nevertheless have a sense of being such a self — are less contentious. There have been a number of arguments from scientific, philosophical, and contemplative quarters for the view that there is no self: at least, no bounded, personalised and persisting centre of agency and ownership from which thoughts and experiences arise.[27]

hours showed less activation. This inverted u-shaped function resembles the learning curve associated with skill acquisition in other domains of expertise, such as language acquisition. The findings support the idea that, after extensive FA meditation training, minimal effort is necessary to sustain attentional focus" (Davidson and Lutz 2008, 173). Davidson and Lutz also cite evidence (with reference to the attentional blink phenomenon) that mindfulness, which aims to cultivate a clear awareness of the field of current experience as it arises moment-to-moment, has lasting effects on the quality of attention outside the context of formal FA meditation practice, such that subjects are "better able to attend moment-to-moment to the stream of stimuli" (2008, 173).

26. In section 5, I offer further speculation on the avenues along which meditative practice might undermine the sense of self.

27. Arguments against the existence of the self have been offered by the authors mentioned note 11. It should be reiterated that the notion of self talked about in this context whose existence is being denied is by no means the only one; as Dan Zahavi (2011, 66–67) and others make clear, there are other notions of self that may well correspond to phenomena that exist. I do however maintain, *contra* Zahavi, that the current notion of self is a central and important one.

I have just provided one such overview as to how the sense of being such a bounded, separate entity could be constructed from patterns of thought and emotion. As for having a sense of self of that description, I have elsewhere offered detailed arguments that we do commonly identify as such a self (alluded to in note 12) as well as some evidence here for the sense of boundedness in particular (see also note 16). With the provisional assumptions now deemed as not unreasonable, and with a clearer understanding of what will meant in ensuing discussion by the terms 'self' and 'sense of self', we can turn to the question of how losing the sense of being a bounded, axiologically salient self, while retaining mental acuity, may be understood in epistemic terms.

2. The Core Proposal: Knowledge-Improvement through Doxastic Integration

Many will be familiar with Frank Jackson's (1986) fictitious Mary, the omniscient colour scientist. Raised from birth in a black-and-white room, Mary learns all the physical theoretical facts that can be known about colour and colour vision. But when the roof of her enclosure opens for the first time to reveal a bright blue sky, Mary's understanding of 'seeing blue' seems dramatically enhanced. As Martine Nida-Rümelin puts it, "there is a strong intuition in favour of the thesis that Mary makes genuine epistemic progress after her release" (2004, 241). Supposing that this intuition is right (and putting aside deeper metaphysical conjectures such as whether this understanding indicates non-physical facts — the original purpose of the thought experiment), there are three main hypotheses which aim to flesh out the idea that Mary now 'knows what it is like' to see blue (even if the locution is, in the end, misleading).[28]

28. For this summary, I draw upon Nida-Rümelin (2010).

Very briefly, the first hypothesis is that Mary gains knowledge of the phenomenal character of blue under a *phenomenal concept* of blueness — a concept whose acquisition requires, as a necessary condition, the direct experience of seeing blue. Once acquired, this phenomenal concept of blue may lead her to have phenomenal beliefs about the colour of various things, such that she can (for example) come to know that the sky appears phenomenally blue to other normal perceivers. Another hypothesis for her improved epistemic status is that she gains new abilities: she can now imagine or recall what a blue sky looks like and she can immediately recognise, without help from scientific instruments, that an object is of a certain colour (defenders of the view hold that 'knowing what blueness is like' amounts to no more than the acquisition of the relevant abilities). This dimension has sometimes been termed *knowledge how*. A third hypothesis is that without acquiring new facts (such as those associated with phenomenal concepts), Mary simply becomes directly *acquainted* with the experience of blueness, which is enough to account for her sense of 'knowing what blueness is like'.

It is not the purpose of this paper to arbitrate on which, if any, of these accounts is correct. But if we accept that colour-scientist Mary epistemically progresses along any of these avenues, then it is reasonable to suppose that Mary Analogue, were she to awaken, would make parallel epistemic progress.[29] Before awakening, she would know all the theory in connection with the proposition 'there is no self' (including the neuropsychology of those Buddhists who have overcome the sense of self). After awakening, she will know what it is like to have no sense of self, whether this be accounted for in terms of her gaining new phenomenal concepts (such that she now knows what arahants phenomenally believe, when they claim that there is no self), or new abilities (such as actually delivering a public speech without fear) or through simply being directly acquainted with the state of having lost the sense of self.

Yet while *some* form of 'knowing what it is like' to have a no sense of self (and indeed a sense of no self) may well be a central component to Mary Analogue's improved epistemic status — according well with Buddhist injunctions that one must 'know reality through direct experience' — there remains, I contend, an important sense in which it fails to capture what is distinctive about Mary Analogue's epistemic improvement. To bring this out, consider the following example. I have never taken mescalin but I've read descriptions of what it is like to take it. Suppose I take it and I am struck with the conviction: '*This is a more correct view of the world.*' Is this thought in fact correct? Well, it depends. If all I mean by 'more correct view of the world' is something localised like 'I now know what it is like to take mescalin' then my conviction may well be right for the sort of reasons outlined above: I may, for example, have gained new phenomenal concepts that bolster what I already knew from reading descriptions of taking mescalin. But suppose I mean something more universal like: 'This experience of taking mescalin is infused with noetic resonance, a *sense of rightness* that exemplifies a profoundly more accurate frame of mind and outlook on the world.'[30] In such a case, there is good reason to doubt the truth of my statement: as with many such drug-induced experiences, I am probably deluded.[31]

Likewise, if post-revelatory Mary were to exclaim that her experience of blueness 'offers a more correct view of the world', her statement

29. If occurring over a series of smaller breakthroughs, might Mary Analogue's final revelation not be as dramatic as that of Mary the colour scientist? The accounts I've read of purported arahants *all* suggest the final breakthrough to be dramatic; but even if this is not so, insofar as Mary Analogue comes to know what it is like to experience a mind completely free from the structures of self, the comparison still stands.

30. By 'sense of rightness' and 'noetic resonance' I mean to convey the feeling of direct intuitive understanding that comes with an 'aha' moment, when things coalesce in a way that appears to make perfect sense.

31. By this example, I do not intend to rule out the possibility that some drug experiences (or neurological traumas) could, in fact, instil (in the radical sense) a 'more correct view of the world', and hence be a way of acquiring so-called insight-knowledge (for a dramatic example of neurological trauma eliciting apparent boundary-dissolving insight, see Jill Bolte Taylor's 2008 TED talk 'My Stroke of Insight'). To my knowledge, however, no such cases that fully match the description of *nibbāna* have been reported.

would be correct up to the point at which she now knows what it is like to experience the sensation of blue, such that she can for instance apply the relevant phenomenal concepts to what she already knew about blueness. But if she were to insist 'No, I mean more than just that: it carries a sense of *rightness*, instilling a profoundly more accurate frame of mind and outlook on the world' she would, as with the mescalin case, be going beyond what she has warrant to claim. But this, precisely, is the kind of description that Buddhist tradition imparts to insight knowledge into no-self. It is not just a matter of claiming 'I now know what it is like to have no sense of self' — on whatever reading. The feeling of overcoming the sense of self is said to carry an additional noetic resonance — a sense of rightness that purports, in connection with having seen through the illusion of self, to instil a profoundly more accurate frame of mind and outlook on the world: a cognitive platform, as it were, from which other chunks of reality, which relate mind to the wider world, are apprehended and understood with far less distortion.

Now what could infuse the experience of losing the sense of self with a *veridical* noetic resonance, making the claim to increased accuracy more legitimate than that of the mescalin or colour-scientist case? A central clue can be found both in Buddhist sources (*e.g.* Bhikkhu Bodhi), and in this further passage by Strawson:

> Consider certain Buddhist philosophers who argue, on a variety of metaphysical grounds, that our natural notion of a persisting individual self is an illusion. Having reached this conclusion, they set themselves a task: that of overcoming the delusion. …They recognize, however, that one cannot simply abolish one's sense of individuality, by some sort of effortless, rationally motivated, self-directed intellectual fiat. Delusions delude, after all; and the ordinary, strong sense of self…is a particularly powerful delusion. They therefore recommend the adoption of a certain practice — that of meditation — the eventual effect of which, they claim, is to cause the delusion to dislimn. [2010, 101–102]

Here is where I propose the difference lies. Unlike in the colour scientist or mescalin case, Mary Analogue's theoretical understanding of the proposition 'there is no self' (and by implication, that she is not such a self) is being coupled with the overcoming of a powerful and pervasive delusion — the delusion that she is a self. Overcoming this delusion imbues her with a genuinely accurate feeling of noetic resonance: of having dispelled a cognitive error — analogous, it is sometimes said, to awakening from a dream. The depth and pervasiveness of the error overcome explains and grounds her feeling that the insight is profound and irreversible, resulting in a correspondingly more accurate mode of cognition.[32] In the mescalin and colour scientist case, there is no error to be overcome, such as a delusion about what blueness looks like; Mary simply learns (or appears to learn) what it is like to experience blueness. While to this extent the colour scientist may well have a more 'correct' view of the world, it is not the sort of correctness that accompanies the overcoming of a cognitive error, nor is it profound in a cognitively pervasive sense. That is why she would not be warranted to make any claims about harbouring a profoundly more accurate state of mind.

From this, we can distinguish two features that characterise epistemic progress towards insight knowledge of no self: the overcoming of a delusion — that is, the dispelling of a cognitive error — and the depth and pervasiveness of the delusion overcome, resulting in a correspondingly more accurate way of viewing the world. In the

32. It is important to note that just as with any intellectual (such as mathematical) insight, a sense of rightness attaching to any purported insight around 'overcoming the delusion of self' will not always be veridical. Buddhist tradition is well aware that practitioners can be mistaken about their claims to insight, which is why the tradition places importance upon other members of the monastic community — especially teachers — to help verify such claims, such as through monitoring reactions to different situations over time. Any display of *taṇhā* (such as through anger, fear, pride, or lust) would for instance be a reliable indication that the practitioner is not awakened.

remainder of this section, I expand Strawson's passage into an analysis of how the Buddhist practitioner, through dispelling the delusion of self, could end up with better quality knowledge of the proposition that there is no self: an analysis which I then apply to other cases (section 4). In the final section (5), I try to account for why we might expect the gaining of insight knowledge, in particular, to result in a substantively more accurate mode of cognition than that which occurs in most other cases of knowledge-improvement.

Strawson's passage carries a number of implications from which we can propose a more detailed epistemic analysis of what it is to overcome the delusion of self. First, there is reference to the self as being both an illusion and delusion — both of which imply that the self does not exist. Illusions occur when an appearance presents the world (to a subject) as having x, when the world does not in reality have x. While often perceptual, illusions can sometimes be cognitive. The illusion of self will be cognitive rather than perceptual: akin to a sense of danger, where the danger does not exist. I take it that by 'delusion' Strawson means a type of false or inaccurate *belief* — minimally, a way one assumes the world to be. In many delusions, the subject will take the content of an illusion to be veridical, thereby assuming the world to carry that feature. Just as an illusion of danger may be assumed, by its subject, to indicate real danger, so the illusion or sense of self is (reflexively) assumed, by the unawakened Buddhist, to indicate a real self. I will put this by saying that the sense of self — a cognitive illusion — is *anchored in* a delusion of self, a belief that the content of the illusion is real.

A second implication pertains to Strawson's claim that Buddhist philosophers (presumably serious practioners) intellectually arrive at their conclusion of there being no self before setting out to overcome the delusion of self. This suggests that the philosophical argument and scientific evidence which justify their reflective belief that there is no self has little effect, by itself, on dislodging their delusional commitment to a self. The delusion that anchors the illusion of self is thus *recalcitrant*, failing to be shifted by standard evidence for its falsity. Contrast this with a non-recalcitrant belief, also based on illusion. Encountering the Muller-Lyer Illusion for the first time, you may innocently assume the two parallel lines to be of unequal length, in accordance with the way they look. You have both an illusion and a delusion that the lines are unequal. Someone places a ruler next to them (or tells you it's an illusion) and you are now correctly convinced that they are of the same length. Although the optical illusion persists, the doxastic anchor has been pulled up. All traces of delusion, both in your actions/affect and in reflective endorsement, have vanished; your initial belief in unequal lines is usurped by the correct integrated belief that they are equal. By contrast, while the introduction of a rationally supported belief that there is no self does indicate a doxastic shift at the level of reflective endorsement, it does *not* pull up the deeper action-based delusion that anchors the persisting illusion of self.

Prior to overcoming the delusion, the recalcitrant belief co-exists with the reflective one, which brings us to a third implication from Strawson's passage: within a given subject there can be the co-presence of contradicting beliefs — an inconsistency of which the subject can be aware. The reflective belief that there is no self will be a component of what we are assuming is the practitioner's *knowledge* that there is no self — propositional knowledge that cohabits with a false action-based belief that there is a self.

A fourth implication suggests that this cognitive state is not epistemically ideal. By overcoming the delusion of self, the subject comes to hold a more correct view of the world, an epistemic improvement signalled by a feeling of direct intuitive understanding I have been calling 'noetic resonance'.[33] But there is a further

33. The noetic resonance seems to imply the emergence of a reflective component that is not reducible to the existing propositional knowledge that there is no self (although it would doxastically integrate with and improve that knowledge). The reflective component would come from the direct intuitive recognition that the self is a delusion. That said, prior intellectual reflection on no self (along with formal meditation) may still contribute to the process of undoing the action-based belief in a self, and hence in gaining complete insight into the reality of no self. I return briefly to this issue in note 42 and in the conclusion.

dimension to the progress, to do with how this shift in what I've been calling 'action-based' belief comes to integrate with the existing propositional knowledge. Before transition, the false action-based belief — from which stems a vast array of *taṇhā*-driven emotions and behaviours — contradicts the reflective belief that is a component of the propositional knowledge that there is no self; after transition, the action-based belief and its attendant attitudes no longer contradict but are in complete conformity with the propositional knowledge. Through becoming consistent, the action-based and reflective beliefs doxastically integrate into one belief (as most beliefs already do). In this respect, I want to claim that one's doxastically integrated knowledge of the proposition 'there is no self' is *better quality knowledge* than that which was had before the transition.

Because the action-based belief (that one is a self) is recalcitrant, the method for attaining this epistemically improved state will not be the conventional route of garnering better philosophical or scientific evidence for the true reflective belief; we can suppose that the best evidence of that sort was recruited in the earlier phase.[34] The route in this case, as Strawson has noted, is rather one of dedicated meditation practice, through which the complex network of *taṇhā*-driven psychological states that feed into the illusion and delusion of self is eventually transformed: the motivations, behaviours, dispositions, affective drives and attitudes.[35] (That the illusion of self must be destroyed in order to dissolve the delusion makes it significantly different from the Muller-Lyer Illusion, where the illusion will remain after the delusion has been dispelled.) The doxastic integration resulting from the final *nibbānic* insight is thus a central respect in which someone with prior theoretical knowledge of her subject matter now has what Strawson terms a 'more correct view of the world' in relation to her subject. We might say that awakened Mary Analogue now *thoroughly knows* that there is no self. In modern idiom, she walks the walk as well as — non-nervously — talking the talk.

34. On a coherentist account of justification (and perhaps a foundationalist account as well), the belief that one is not a self may, after loss of the false and contradicting belief one is a self, become better justified in virtue of the set of beliefs having greater overall coherence. It is not clear, however, that this should be described as a case where the subject has better evidence for the belief, as the action-based belief was recognised as false to begin with. The related question of what sort of evidential and justificatory role might be played by a veridical experience of having overcome the illusion of self, is, I think, an important one that will have to be deferred to another occasion. I return to it briefly in the conclusion.

35. Although the final insight will itself have a reflective component — for more on this see notes 33, 42, and the conclusion.

3. Defending supporting claims of the core proposal

The above analysis, which proposes a core epistemic component for the gaining of insight knowledge into no self (that of knowledge-improvement through doxastic integration), relies upon the truth of at least three contentious claims. To reiterate, these are: first, that the sense of self is anchored in a (false) action-based belief; second, that this false belief that one is a self can co-exist with a contradicting reflective belief that there is no such self (a component of the subject's propositional knowledge); and third, replacement of the false, contradicting belief with a true, consistent (action-based) belief will, via doxastic integration, improve the quality of the propositional knowledge that there is no self. What follows are some suggestions on how each of these claims may be defended.

3.1 The sense of self is doxastically anchored.

Despite the self being commonly called 'a delusion' in Buddhist literature, many will deny that the sense of such a self — whether the self exists or not — can be anchored in a belief that one is such a self. Most centrally, the sense of self's reflexive and elusive mode of presentation will mean that its content is not immediately accessible to introspection and therefore not amenable, via that channel, to reflective endorsement.[36] To believe that P, the objection goes, one has to

36. If one does reflectively endorse the proposition 'I am a self', it will be the result of an exercise which objectively reflects upon the inferred content of self and affirms its existence. This does not make sense of self *in itself* doxastic.

be readily disposed to consciously access and reflectively judge that P, where the disposition to judge that P is part of a broader disposition-base to follow various norms of rationality. Such norms include revising the belief in the face of changes to our all-things-considered evidence, being prepared to bet high stakes on P being true, and believing propositions that follow obviously from P. The sense of being a self (for example, its potential recalcitrance in face of countervailing evidence) disobeys at least some of these norms. Those defending a doxastic account of the self-sense can respond that this rationalistic 'judgement-based' conception of belief — what I have so far alluded to as 'reflective' belief — is not the only one available; philosophical tradition also recognises what I have been calling an 'action-based' approach, where beliefs can be ascribed on the basis of non-reflective criteria such as observable patterns of emotions and behaviours. Such criteria, which include those exemplified in belief-desire analyses of behaviour,[37] may permit the ascription of belief to non-rational creatures such as animals. If the sense of self is doxastically anchored, then it will be anchored in the sort of belief that is ascribed along an action-based rather than judgement-based avenue.[38]

37. On the belief-desire analysis (characterised, but not endorsed by Velleman), a belief that P "dispose[s] the subject to behave in certain ways that would promote the satisfaction of his desires if its content [P] were true" (2000, 255). The term 'action-based' also includes reference to patterns of emotion, *e.g.*, if S desires that P, then coming to believe that P will elicit positive emotion, and coming to believe not-P will elicit negative emotion (Zimmerman 2007, 64). Can these standard action-based criteria be used to indicate the sort of reflexive, action-based belief in the self's existence that would be instantiated by harbouring a sense of self? Not by directly substituting 'the self exists' for P. While I shortly discuss the implications of this in 3.2, I will for now suppose that action-based criteria pertaining to manifestations of *taṇhā* apply.

38. While the terms 'judgement-based' and 'action-based' are my own (and are given a detailed formulation in Albahari 2014), variants of these two approaches to belief-ascription have been described under different names. For example, H.H. Price, as noted in Gendler (2008a, 638n), attributes what he calls the 'traditional' (*i.e.* judgement-based) view to Descartes, Hume, Spinoza, Cardinal Newman and Cook Wilson. Insofar as they "privilege evidence accessible from the third-person perspective", Zimmerman (2007, 71, 72–73) attributes versions of a "third-personal" (action-based) position to Williamson (2000), Smith (1994), Stalnaker (1984), Davidson (1984), and Dennett (1987). While third-personal or action-based approaches (such as an interpretive stance) can allow the display of 'not-P-ish' behavioural and emotional patterns to trump contrary displays of reflectively endorsed belief that P (such that S is ascribed a belief that not-P), we should not assume that they all do. In cases of conflict between what is reflectively endorsed and how one acts, S might not be ascribed any belief at all, or might be ascribed both a belief that P and a belief that not-P. My analysis of insight knowledge depends on taking the latter 'disjunctive' approach to belief-ascription. More on this soon.

Assuming the viability of an action-based approach to belief ascription (an assumption that would need further defence), have we reason to suppose that the sense of the self could actually be anchored in such a belief? Here is an initial reason. The sense of the self can be described as a conscious impression of *being* a self, namely, of being an elusive entity with such features as boundedness, agency, and axiological salience. As the self (qua subject) eludes direct introspection, these features will be indeed be ascribed largely on basis of behavioural and emotional patterns, and I have already provided some examples (for instance in note 12) of how such features may be inferred. That we identify as being a bounded entity in particular is most broadly evidenced, as we saw in section 1, through manifestations of *taṇhā*: the spectrum of desire-driven thoughts, emotions and behaviours that tacitly assume a *salient thing* on behalf of which the desires and emotions are felt and the actions carried out. Now barring beliefs with tautological content, a necessary component of any belief, whether judgement- or action-based, is that it contains truth-apt content that can be in error — and this indeed is being claimed about the sense of self. For if the self (we have a sense of being) turns out not to exist, then our sense of self will be perpetuating some kind of psychological error, or, as is commonly said, an illusion. We will not in reality be the sort of thing that we unwittingly take ourselves to be. From this, it may be tempting to infer that we commonly do harbour a (false) action-based belief that we are a self; the illusion is anchored in a delusion.

Matters are not, however, so straightforward. In a paper which argues that a subjective sense of libertarian freewill does not entail a belief in such freewill, Richard Double (1991) invites us to consider,

amongst other examples, the Muller-Lyer Illusion. While the content of the perceptual illusion misrepresents reality, so contains truth-apt content that allows it to be in error, this "pre-emptory belief" (as he calls it) in no way implies that we harbour a genuine belief about the uneven length of the lines. As my earlier example showed, countervailing evidence can make us instantly switch from a state of believing the content of the illusion to disbelieving it — where 'belief' is ascribed on action-based as well as judgement-based criteria. (Behaviours around attempting to re-design one's house using 'Muller-Lyer technology' will, after initial disappointment, be shifted). Similarly, the objection goes, the cognitive illusion of having libertarian freewill — or of being a self which has it — may beguile the philosophically naïve, but those who know better will not be taken in by it. Double's objection is actually double-barrelled. First, it suggests that an impression or sense or illusion of x does not automatically guarantee a belief that x, so there is no pressing reason to suppose that the sense of self is actually evidence of a belief in a self. Second, it suggests that if there *is* any initial belief in the self (whether action- or judgment-based) then will be revised by countervailing evidence, leaving us with no reason to suppose that an action-based belief in the self would persist in the case of those who, on judgment-based criteria, come to believe there is no self.

In response to the first point within Double's objection, we can note is that while a perceptual or cognitive impression of x does not guarantee a belief that x, it is nevertheless true that without countervailing evidence, an impression of x will usually be unquestioningly assumed to indicate a real x, allowing us to ascribe to the subject an action-based belief that x. The default position, after all, is to believe and act upon the deliverances of one's senses; a systematic scepticism is not how we have evolved to engage with the world. Presented with appearances of grass, trees and people, we take for granted their veracity.[39] Just as there is cognitive pressure to accept deliverances of the perceptual senses, such pressure commonly exerts itself in the case of conscious impressions, particularly if the impression is powerful and persistent. We will have heard about the schizophrenic who, in the absence of knowledge about her condition, thoroughly buys into the content of such impressions as having her thoughts monitored by secret police, etc. — we infer it from the paranoid, fearful behaviour. The conscious impression of being a self will not only be powerful and persistent, but will tend to be regarded as neither abnormal nor pathological nor false. Hence, at least in cases with no countervailing evidence at hand, it seems at this stage reasonable to infer that the conscious impression of being a self, just like our sense-impressions, is anchored in a powerful default action-based *belief* in the veracity of that impression.

39. We can draw a connection here with Wittgenstein's notion of a framework belief, discussed in section 3.2.

3.2 Delusion of self co-exists with knowledge that there is no self

If we provisionally grant that the sense of self is normally anchored in a false action-based belief that one is a self, can the delusion be held in conjunction with an opposing judgment-based belief — indeed knowledge — that one is not such a self? Here we are faced with what appears to be a two-horned dilemma. The first horn pertains to the second point within Double's objection above. He would contend that a judgement-based conviction that one is not such a self (elicited by countervailing evidence) will, just as in the case of the Muller-Lyer Illusion, serve to replace any opposing, action-based belief that one is a self with a consistent and integrated action-and-judgement-based belief that one is not such a self. If Double is correct here, then my proposed analysis of insight knowledge must fail, as it depends upon the delusion of self being recalcitrant in the face of evidence that supports an opposing, judgment-based belief. But if the other hand Double is not correct and the so-called delusion of self persists in the face of its apparently opposing conviction, then we have reason to suppose that the sense of self is not actually anchored in a belief at all, but in something pre-doxastic, such as an alief or default psychological architecture. In addressing this second horn of the dilemma, we revisit the

assumed viability of an action-based belief as best explanation for the 'delusion' of self. If the sense of self is not after all anchored in a belief, then my analysis must fail.

To the first horn: I think that Double's point about the Muller-Lyer Illusion does not carry over to that of the self (including a self in its capacity of harbouring libertarian free will). While discovering the real status of the optical illusion would involve a discernable shift in our underlying emotional and behavioural attitudes towards the phenomenon, so that we no longer take it seriously, becoming convinced (on judgement-based criteria) that there is no self is *not* likely to elicit an analogous shift in manifestations of *taṇhā*, such that we take the self less seriously.[40] As Strawson implies, there is little evidence to suggest that philosophers and Buddhist practitioners who become intellectually convinced that there is no self will be any less emotionally invested in protecting the imagined boundaries of their selves than they were before. Indeed, there is likely to be no discernable difference in the levels of *taṇhā* between those who reflectively endorse the existence of self and those who deny it. Mary Analogue's fear at public speaking is not merely the stuff of fiction.

Why might this be so? First, those advocating the reality of action-based beliefs often note their connection with deeply ingrained modes of conditioned or instinctual response to the environment that are at best slow to respond to contrary evidence, and in section 4, I consider several further candidates for such belief.[41] The beliefs may be especially stubborn if the mode of response has evolved to aid survival, as Damasio supposes in the case of the self. A further reason for recalcitrance in the case of the self, I surmise, lies in the structure of the subject/object division. Lying on the subject side of the division, the reflexive content of the self systematically eludes direct observation, and so cannot be the immediate target of its own observational scrutiny. Revelation of its illusory status (and the subsequent revision of belief) could thus never happen via the same sort of direct observational channels that occur, say, with the Muller-Lyer Illusion; there's no equivalent of placing a ruler next to the lines. It stands to reason, then, that an overturning of the default action-based belief that one is a self will have to come about through methods (such as meditation) that erode the very psychological structures that subtend the sense of being a self.[42] This would account for the lack of parallel, mentioned earlier, between the illusion of self and most other perceptual (and likely, cognitive) illusions. Whereas in most cases, the illusion can or does persist after the delusion has been dispelled, this is not the case with the sense of self. Being reflexive, the impression that one is a self cannot come apart from the belief that one is a self; *the* taṇhā*-driven cognitions that constitute the sense of self double as the vehicle through which the self is assumed*. The cognitive relation between sense of self and its doxastic anchor is thus a very close one. For these sort of reasons, we should expect such an action-based belief to remain recalcitrant in the face of opposing theoretically-based evidence.

Yet the very considerations that speak in favour of the 'belief' in self being recalcitrant work against its being doxastically anchored, taking us to the second horn of the dilemma. Could the deep recalcitrance, along with the lack of parallel with other illusions, indicate that the sense of self (even without the co-presence of a conflicting belief) is not doxastically anchored at all, but is rooted in something entirely more primitive and pre-doxastic, such as an alief or basic psychological architecture? Damasio and Panksepp both regard the impression of

40. Recall our supposition that emotional and behavioural manifestations of *taṇhā* comprise the relevant sort of action-based evidence for believing one is a self.

41. While I shortly consider an alternative hypothesis of alief, a good discussion of dual-process cognition, which offers a psychological explanation for the recalcitrance of various beliefs that are "typically slow to form and change", can be found in Keith Frankish (2009).

42. Through such processes, there would come a stage at which the sense of self is eroded enough to permit the final purported burst of insight, upon which the delusion of self is seen through and comprehended in the direct and intuitive way that I've been calling 'noetic resonance'. While such experience does suggest a mode of direct observation, I surmise that it would transcend normal constraints of subject versus object, so that ordinary notions of elusiveness (which presuppose this division) no longer apply. This is a topic for further investigation.

a self to arise at a primitive neurological level, involving circuitry that combines basic sensory-motor with affective inputs. Taking their accounts seriously, as I have done, seems to favour a pre-doxastic analysis. I now consider the pre-doxastic hypotheses that the sense of self is anchored in (a) an alief, and then, (b) basic psychological architecture, such as a model the brain creates to allow the organism to function in the world.

To the alief hypothesis. Aliefs (the term was recently invented by T.S. Gendler, 2008a, 2008b) are defined as primitive, pre-doxastic, pre-rational, clusters of reaction to apparent stimuli, that associate representational content (*e.g.* a precarious-looking height) with affective reactions (*e.g.* feelings of fear) and behavioural proclivities (*e.g.* an urge to step away). Gendler introduces the notion of alief to provide a unifying explanation for those puzzling phenomena where behavioural and affective tendencies persist in spite of rational beliefs to the contrary: think of the common reaction (butterflies, shaking) to stepping on the glass Skywalk above the Grand Canyon that we rationally know is safe. While the norm for belief, she says, is to be rationally responsive to all-things-considered evidence, making it appropriate to criticise as irrational if stubbornly persisting (such as a belief that one is a better driver than one is), alief is not governed by such norms of rationality (2008b, 570). While aliefs may be seen as undesirable (*e.g.* in cases of racism), their inherent unresponsiveness to all-things-considered evidence makes it inappropriate to deem them irrational. Any changes to formations of alief must occur gradually, through processes of association and conditioning. From everything said so far, it would seem that the deeply recalcitrant, survival-promoting 'assumption' of being a self, which would automatically and reflexively associate the content <bounded self> with feelings of emotional attachment and behavioural proclivities to protect, is a prime candidate for alief.

Alief is a controversial cognitive category; some argue, for instance, that it deflates into belief. My own take on it, that I develop at length elsewhere, is that alief is an independent cognitive category that (contra Gendler) is not the main unifying explanation for all motor/affective tendencies that clash with judgement-based belief (Albahari, 2014). In particular, alief does not primarily explain those instances of discordant behaviour and emotion that fit standard criteria for action-based belief; action-based belief explains those.[43] Alief offers the primary explanation for cases where the motor-affective tendencies are better described as modulating the dominant, judgement-concordant affective-behavioural arc with opposing tendencies.[44] Compare the typical behaviours and emotions of a height-phobic and a non-phobic who find themselves on the glass Skywalk. While both have vertigo and rationally judge the platform to be safe, only the phobic's behaviour and emotion fits standard action-based criteria of belief-ascription — she feels her life to be in danger as she desperately tries to leave the platform. She has a (contradicting) belief that the platform is unsafe.[45] The non-phobic's butterflies and hesitancy in stepping, by contrast, are caused by aliefs that serve to modulate her overarching, action-and-judgement-based belief (manifested through gut-level feelings of safety and stepping on the platform) with opposing tendencies.[46]

43. For evolutionary reasons discussed in that paper, I reject Gendler's contention that belief can only be ascribed on judgement-based criteria, as guided by norms of rationality.

44. I say 'primary explanation' as I hold that alief has some explanatory role to play in every case where there is that triplet of 'RAB' association, including those of contradicting beliefs; it is just that in these latter cases alief is not the main cause of the discordant reaction. I say more about complementary roles of alief and belief in note 46.

45. I recognise that the notion of having contradicting beliefs is contentious; more on this soon.

46. In that paper I propose that these criteria for distinguishing alief from belief (in the capacity of explaining discordant tendencies) are aptly grounded in their different and complementary evolutionary roles. I argue that the evolutionary role of belief — in keeping with the common belief-desire platitudes — is to guide and execute behaviour (in conjunction with prevailing desires). The function of alief is associative rather than executive; its role is to speed up reaction time by associating representations with affective and motor responses, so that when feeding into the action-guiding belief circuitry, the actions are carried out much faster. Hence aliefs will always manifest as *partial* (never fully carried out) reactions that speedily associate representational,

If my account is correct, then the case of the self does not fit the profile of a mere alief, for it resembles the case of the phobic more than that of the hesitant stepper.[47] Consider the typical self-denier's thought that something unpleasant is about to befall them. Their *taṇhā*-influenced emotions and behaviours (taken, so far, as the most likely action-based criteria for affirming the existence of self) are not aptly described as merely modulating a dominant action-and-judgement-based belief in no self with self-like tendencies. *Taṇhā* drives their emotions of trepidation and behaviours of avoidance. As noted earlier, their level of *taṇhā* is likely to be no less pronounced than in those who reflectively endorse the self's existence. So if not anchored in merely an alief, is the sense of self anchored in an action-based belief? Here we now face a different problem; the 'reflexive belief' in the self's existence does not conform to standard action-based criteria.

Let us revisit the criteria at hand. On the belief-desire analysis, a belief that P "dispose[s] the subject to behave in certain ways that would promote the satisfaction of his desires if its content [P] were true" (Velleman, 2000, 255). On the patterns of emotion analysis: If S desires that P, then coming to believe that P will elicit positive emotion, and coming to believe not-P will elicit negative emotion (Zimmerman, 2007, 64). If we substitute 'the self exists' for P, we quickly see that neither formula applies. For a start, the existence of self is not something that from an action-based perspective we come to believe, which would in turn cause various emotional reactions. Indeed, the standard action-based criteria (by which a subject is disposed to behave in ways to promote the satisfaction of desire — becoming happier the desires are satisfied and less happy if they are not) presuppose the subject to be in the grip of *taṇhā*. So far I have been taking *taṇhā*-driven emotions and behaviour to constitute unique action-based evidence for the reflexive belief that one is a bounded self. But could this anomaly in the so-called action-based criteria — by which an 'assumption' of self is presupposed in the very having of an ordinary action-based belief — suggest another hypothesis? Could the sense of self be anchored not in a belief that one is a self, but in basic, pre-doxastic psychological architecture or a 'self-model' that grounds much of our ordinary psychological practices and behaviours?

I propose that something in between is correct: that the sense of self, while indeed anchored in a deep psychological structure of the sort described by Damasio and Panksepp, is nevertheless anchored in a structure that is doxastic at its core. The vast edifice of neurologically-based affective and motor proclivities that comprise this structure serves as the vehicle that bears the unifying doxastic content <I am a bounded self>. But what kind of action-based belief could it be? While a full defence is not possible here, I suggest that what I've been calling the 'assumption' of self closely fits the profile of a framework belief (or hinge proposition) along the lines described by Wittgenstein in On Certainty (1969) and developed by later thinkers such as Lisa Bortolloti (2010). Framework beliefs are axiomatic assumptions that play a foundational role in the formation of other beliefs; they are central to our worldview. Rather than being objects of overt knowledge or belief within the framework, they are appealed to or assumed when justifying or forming other items of knowledge and belief. Framework beliefs are typically taken completely for granted. As Bortolotti puts it: "The commitment to a framework proposition is pervasive and manifested in many instances of behaviour, although the belief remains in the background and may never be explicitly reported or justified" (2010, 192). They are likely to be discovered, rather than explicitly learnt:

> I do not explicitly learn the propositions that stand fast for me. I can *discover* them subsequently like the axis around which a body rotates. This axis is not fixed in the sense

motor and affective content — most noticeably in belief-discordant cases, but also in the vitally important belief-concordant cases, such as when hastily stepping away from a moving car (Albahari 2014).

47. The examples of phobia, superstition, and clinical delusion, as I describe them in the following section, will also fit the profile of belief rather than alief, although a doxastic diagnosis will not uniformly apply to *all* such phenomena within those categories (*e.g.* clinical delusions).

> that anything holds it fast, but the movement around it determines its immobility. [Wittgenstein, 1969, §152]

Framework beliefs tend to be integrated within one's belief system and shared by a linguistic community. Common examples given of framework beliefs are those that express assumptions about the reality of an external world, such as 'I have two hands', or 'the world existed before I was born'. I suggest, then, that our reflexive commitment to the existence of a bounded self (along with our supposition that others are such a self) is one of the most foundational framework beliefs we have. Revolving around the axis of this belief will be an enormous network of derivative beliefs, psychological and linguistic practices — including, as noted, the standard criteria for the ascription of ordinary action-based beliefs. On this model, all thoughts and beliefs that express personal apprehensions about the past and future, all expressions of preference and aversion, all thoughts of identification and claims to personal ownership ('me' and 'mine') — in oneself and towards others — presuppose an action-based framework belief that one is a bounded self.

There is debate over whether framework beliefs can be meaningfully denied, revised or justified, or whether they are true, or indeed to be properly regarded as beliefs or components of knowledge at all (Wittgenstein expressed doubt over this). Without entering into this debate, I note my agreement with Bortolloti:

> Wittgenstein did stress that there are some beliefs that are more basic and less open to revision than others. But he fully recognised that there is no sharp distinction between the hard rock at the bottom of the river and the sand: both are subject to change. We would be mistaken if we took his description of the bedrock to mean that some beliefs *cannot* receive justification or can *never* be revised. [2010, 196]

Viewing the 'assumption' of self as an action-based framework belief that can possibly be revised is illuminating, as it throws into relief the utter enormity of what would be entailed by its eventual revision.[48] The initial assertion of an opposing judgement-based belief 'there is no self' now appears as a mere chipping at the tip of an iceberg, a tiny shadow of contradiction. With the disintegration of the vast edifice of cognitions that express self-assuming preferences, aversions, anxieties, identities, etc., and their replacement by a set of cognitions consistent with there being no self, would come a radical alteration of one's entire way of thinking and living — exactly as described in Buddhist texts.[49] I return to this theme in Part 5.

We must still establish the general claim that a subject can have contradicting beliefs, and be aware, moreover, of having them. While this will not be a claim that I defend exhaustively here, it can be noted that those adopting a purely judgment-based approach to belief-ascription are most likely to resist such a possibility. For being aware of having contradicting, reflectively endorsed beliefs may well commit the subject to the assertion of problematic Moore-paradoxical sentences, such as: 'There is no self, but I believe that there is'. Elsewhere (Albahari, 2014), I defend what I call a 'disjunctive' approach to belief-ascription, which allows belief to be ascribed to a subject on either

48. While I do not hold the potential for revision to be a necessary condition of belief (as I think that animals can have non-revisable action-based beliefs), it is worth noting (on the issue of whether the sense of self is doxastically anchored) that those who do uphold this criterion will put more store by whether the nibbānic hypothesis holds. Should it turn out that Damasio and Panksepp are correct and the sense of self cannot be overthrown (without serious pathology) then this would count against the doxastic reading and in favour of the (mere) psychological architecture model. But if the assumption of self can be revised — especially in a manner that results in the practitioner understanding the revision via the experience of noetic resonance — the plausibility of the doxastic reading increases.

49. This carries the interesting implication that standard action-based criteria for the ascription of ordinary beliefs would fail to apply to the arahant. And while arahants would continue to use I-sentences, they would not be used to express personal preferences, aversions, identities etc. The 'I' would indeed only serve to function, in their minds, as a convenient designator for their perspectivally owned bundle of thoughts, perceptions, feelings, and so forth.

judgment-based or action-based criteria — an account that is essential to my current analysis. On this approach, I see nothing particularly problematic about subjects having a reflectively endorsed judgment-based belief that contradicts a disavowed action-based belief, inferred and self-ascribed through observation of their own emotion and behaviour.[50]

3.3 Doxastic integration improves the quality of propositional knowledge

Suppose we accept that a false action-based belief, that one is a self, can be usurped by an action-based belief that integrates with the doxastic component of the existing knowledge that there is no self. Put comparatively: at t1 there is a false action-based belief that not-P, together with propositional knowledge whose doxastic component is a judgement-based belief that P; at t2 there is propositional knowledge whose doxastic component is an integrated judgement-and-action-based belief that P. Does the doxastically integrated knowledge at t2 (that there is no self) improve upon the quality of that knowledge at t1? The idea that the quality of propositional knowledge can vary has been famously defended by Stephen Hetherington (2001, 2006), who argues that such knowledge is not an absolute 'all-or-nothing' affair, but can vary in quality along the justificatory axis. On his version of this 'gradualist' account, knowledge with infallible justification is 'extraordinary knowledge', while knowledge with fallible justification (with possibilities, such as dreaming, that could rule out the truth of P) is lesser-quality 'ordinary knowledge'.

Adam Leite objects that this fails to sit with our ordinary judgement, as "we don't judge it an improvement to seek out evidence against possibilities that we all agree have no reason in their favour" (2006, 323). A possible defence against his idea that the notion of knowledge-improvement jars with ordinary judgement is to allow, as Richard Kitchener (2003) suggests, that knowledge might also improve along doxastic (not just justificatory) lines. While Kitchener views doxastic improvement in terms of degrees of subjective confidence in one's belief, my account would also view doxastic improvement in terms of the integration of action-based with the true judgement-based belief. Far from jarring with our ordinary judgement to suppose that one's knowledge might improve along such axes, it fits well with the common intuition that if a subject S correctly judges that P, and yet otherwise lives and behaves as if P were not true, there is something irrational about S; and that, conversely, S's bringing her unreflective modes of thought and behaviour into line with her considered beliefs makes her more rational. We sometimes say that there is a loss of cognitive dissonance. Through ascribing to S *better-quality knowledge*, for the doxastic reasons outlined, my analysis serves to place this loss of cognitive dissonance within a more concise epistemic framework.[51] Of course, further work will need doing to fully defend a gradualist account against other objections from 'absolutist' quarters which, after all, is the prevailing view.

4. Wider application

Jim, a schizophrenic, is gripped with a sense of being monitored by aliens – a clinical delusion that permeates many facets of his life. He deactivates his smoke alarm for fear of it being a 'listening-device', he keeps his curtains drawn at all times, his heart palpitating at unexpected noise. Yet Jim has awareness of this situation: he knows, intellectually, that he is clinically deluded and that there are no such aliens. A course of medication helps dispel the delusion, such that his deeper emotions, motivations, thoughts, and behaviours no longer

50. The existence of a disavowed action-based belief in one's own cognitive set would perhaps be most naturally expressed through saying 'I have a belief that I am a self' rather than 'I believe that I am a self'. For recent accounts that recommend a disjunctive approach to belief-ascription (allowing for beliefs that contradict each other), see Sommers (2009), Frankish (2009), Gertler (2011), and Albahari (2014). For recent accounts that would oppose a disjunctive approach to belief-ascription in the face of apparently contradicting beliefs, see Schwitzgebel (2010), Zimmerman (2007), Gendler (2008a, 2008b), and Kriegel (2012).

51. As described in note 34, the loss of false and inconsistent belief may also incur improvement along justificatory lines, for reasons of making the set of beliefs more coherent. The reply to Leite would still stand.

appear to jar with his reflective knowledge. Heather has a phobia of feathers. While intellectually judging that feathers are not dangerous, her daily sightings of birds make her life a living hell, the mere sight of a gull causing her to flee the beach, abandoning her young child. Exposure therapy gradually alters her mindset, such that Heather suddenly 'realises', one day, 'feathers are not dangerous!' Of course she already knew that feathers are not dangerous; the difference, this time, is that her motivations, emotions, and behaviours no longer seem to conflict, but appear congruent with her intellectual knowledge of that fact. Kla Han, a young physics student raised in a Thai village, is attending a prestigious Western university. Kla Han intellectually knows that ghosts do not exist, and yet a deep culturally-instilled fear of ghosts prevents him from taking a convenient shortcut home across the graveyard at night. Peer pressure eventually drives Kla Han to a hypnotist who helps him to overcome his fear, such that his emotions and behaviours now seem congruent with his existing intellectual knowledge.

The cases above describe various emotions and behaviours as appearing to conflict and then align with the relevant intellectual knowledge. Applying my analysis of knowledge-improvement to such cases will explain the appearance of conflict and its dissolution in literal, doxastic terms: action-based beliefs that are ascribed on the basis of those patterns in the emotions and behaviours.[52] There will be an initial contradiction between the false action-based belief and the judgement-based knowledge of the case at hand, followed by a supplanting of the false, inconsistent belief (through methods that don't aim at improving evidence for the existing knowledge-claim), with a correct action-based belief that integrates with the judgement-based doxastic component of the existing propositional knowledge. Any 'sense of rightness' will be grounded in the event of having genuinely overcome a delusion, such that the individual in question really does exemplify a more consistent and accurate frame of mind with reference to the proposition at hand (*e.g.* 'feathers are not dangerous'). Knowing that proposition more consistently and thoroughly than before, the subject's knowledge will be of better overall quality.

The account should also fit those cases which, like that of the self, are not pathological, but nevertheless involve a clash between a well-argued philosophically endorsed belief (if the belief is true) and a mode of living and thinking. A classic example is that of determinists who intellectually believe, let us suppose on good grounds, that there is no such thing as libertarian free will, while at the same time agonising over decisions in a way that would suggest that they do subscribe to such freewill. On this, Van Inwagen writes:

> …to reject [libertarian] free will [as the determinist does] is to condemn oneself to a life of perpetual logical inconsistency. Anyone who rejects free will adopts a general theory about human beings that he contradicts with every deliberate word and act. [1983, 160][53]

If we accept Van Inwagen's analysis, then if determinism *is* correct it seems natural to conclude, in accordance with my account, that one's knowledge of determinism would improve through replacement of the contradicting, false belief in libertarian free will with a doxastically integrated belief in determinism. Another common example is that of the moral anti-realist who, despite arguing against the existence of moral facts (believing all moral judgements to be literally false), lives a highly ethical life. He feels strong moral sentiments, makes passionate moral judgements and attempts to act in accordance with them.

52. According to my earlier-cited criteria for distinguishing the ascription of alief from belief, the examples *as I've described them here* should count as cases of action-based belief rather than simply alief. To re-iterate: I do not regard all cases of judgement-discordant tendency to be doxastic. Many cases will involve just alief. The doxastic status of clinical delusions, moreover, is a topic unto itself that has spawned much debate, including whether they could be framework beliefs. Should any of my examples qualify as instances of framework belief, it will be suggested in the following section that they are not as foundational as the belief that one is a self.

53. It is to Van Inwagen's account that Double responds with his argument against the feeling of free will being doxastic.

Each act of moral reflection seems to contradict his philosophically informed belief that moral facts (or properties) do not exist, creating an aura of epistemic irrationality. If moral anti-realism were correct, then the quality of the moral anti-realists' knowledge (although perhaps not their lives!) would improve if they quit feeling, thinking, and behaving as if moral properties existed.

Returning to the case of the self, Hume famously wrote:

> However at one instant we may consider the related succession [of ideas] as variable or interrupted, we are sure the next to ascribe to it a perfect identity, and regard it as invariable and uninterrupted. Our propensity to this mistake is so great from the resemblance above-mention'd, that we fall into it before we are aware; and tho' we incessantly correct ourselves by reflexion, and return to a more accurate method of thinking, yet we cannot long sustain our philosophy, or take off this biass from the imagination. [Book 1 Sec 6, Part IV, 1739–40/1978, 254]

Had Hume removed the bias from his imagination by doxastically integrating action- with judgement based beliefs, he may well have come round to a more accurate method of thinking. Despite the fame of his passage, however, there is little reason to suppose that Hume ever succeeded in doing this properly. For the idea of "correct[ing] ourselves by reflection" such that we can "return to a more accurate method of thinking", makes it sound as if we can puncture the delusion of self simply by concentrating hard enough on our philosophical beliefs. For reasons already cited, this will not suffice.[54] Buddhist traditions unilaterally emphasise the need for long-term meditation practice to destroy the sense of self and its anchoring delusion. So how does this connect with the idea that its eventual dissolution would be correspondingly insightful in a way that elimination of an ordinary pathological delusion or phobia would not be? This brings us to the second part of my analysis of what it is to gain insight knowledge. I've already suggested that overturning the delusion of self would involve the revision of a most basic framework belief, but it remains to be seen how overcoming such a delusion could have a profoundly transformative effect on the *accuracy* of cognition. To this final section we now turn.

54. This is not to deny that sustained philosophical reflection (either on its own or with meditation) could make a notable difference to the sense of self, such as in Parfit's description of his own case (1984, 281). It is rather to deny that philosophical reflection *alone* could entirely puncture the delusion of self.

5. Removing veils of cognitive distortion

In the earlier-cited passage, Bhikkhu Bodhi writes:

> ...wisdom removes the veils of distortion, enabling us to see phenomena in their fundamental mode of being with the vivacity of direct perception. [1994, 56]

In this section, I offer some speculations on how 'veils of distortion' pertaining to the illusion of self could manifest, such that their removal (via meditative techniques) would result in a substantively more accurate mode of cognition. My goal, once again, is not so much to offer a full elaboration or defence of the proposal, as to indicate a plausible direction that further research on the topic may take.

As a first pass, I propose that overcoming the delusion of self would differ in degree to that of overcoming any clinical delusion, phobia or superstition by simply occurring at a more fundamental psychological level. For consider: even after overcoming a clinical delusion or phobia, the sense of self remains intact; one still identifies as the subject of the delusion overcome (*e.g.* 'no aliens are watching *me*'). That the delusion of self persists through normal states of mind would indicate that its undoing is a far greater undertaking than that of any clinical delusion — well in keeping with the hypothesis that sense of self is anchored in one of our most basic framework beliefs.

There is a further reason why overturning the delusion of self would have a deeply pervasive effect on cognition. The reflexive mode of presentation — by which features of the self appear to qualify the

observing subject of experience — would mean that any illusory aspects of the self will serve to distort, as it were, the very *lens of cognition* through which all aspects of the world (including one's thoughts and experiences) are apprehended. It is likely, then, that dissolution of such illusory features would have a globally transformative effect upon the accuracy of cognition. I now offer a hypothesis on how these reflexive veils of distortion, and their subsequent dissolution, could affect the accuracy of cognition.[55]

If we accept Damasio's account of the construction of the personalised self/other boundary, outlined in section 1, we can point to at least three distinct avenues along which the illusion of self may be perpetuating distortion. First, attention (often absorbed in thoughts about one's past or future) will be persistently skewed towards what is of perceived relevance to the assumed self. This would suggest that the attention-and-action-guiding thoughts (*saṅkhāra*) and perceptions (*saññā*) are systematically selective rather than impartial, with many aspects within one's current purview going unnoticed, and other aspects emphasised (think of Mary Analogue noticing mainly the famous philosophers). Second, memories of objects and situations will be bound up with the "obligate emotional reaction" to those objects, as one repeatedly reacts to the pleasant or unpleasant valences (*vedanā*) with preference or aversion (*taṇhā*). This will colour our perceptions (*saññā*) of the world, such that we consciously or unconsciously judge its contents to be inherently value-laden: 'good', 'bad' or 'irrelevant', for example. Independently of such judgements, the contents are likely to be neutral. Third, and perhaps most fundamentally, the reflexive sense of being a separate, bounded, axiologically salient entity (on behalf of which desire-satisfaction is sought) — an entity that is the content of a deep cognitive illusion — will obscure apprehension of what would otherwise be a far more dynamic, fluid, and permeable relation between the world and one's psycho-physical existence.

This analysis provides a specific basis from which to conjecture on just how losing the sense of bounded self, through overcoming the delusion, could substantively increase the accuracy of cognition. Recall that Buddhism holds it possible to sever the link between pleasant or unpleasant sensation (*vedanā*) and one's mental reactions of preference and aversion to those sensations (*taṇhā*). For the severance to occur, the practitioner must first be made keenly aware of those reactions as they unfold in real time. The practices of meditation (*samādhi*) and mindfulness (*sati*) aim to significantly improve the level and scope of attention paid to the unfolding phenomena, allowing the practitioner to nip in the bud *taṇhā*-driven reactions as they arise (such as by refusing to proliferate on I-thoughts in response to pleasant or unpleasant sensations).[56] This gradual elimination of self-assuming cognitions, culminating in the final *nibbānic* insight, incurs the reciprocal loss of cognitive distortion along at least the three lines mentioned above.

First, to modify a phrase from Hume, attention would no longer be a slave to the passions but would be far more impartial, such that one is disposed to notice, within one's current purview, more phenomena and with greater accuracy as they unfold in real time. Mental acuity would be greatly enhanced by the high quality of attention that would be cultivated in the process of freeing it from the deeply rooted impulse to attend to and satisfy self-related cognitions. Second, there would be significant reduction or elimination in the projection of self-related values onto situations and objects, such that the world

55. It might be objected that various clinical delusions and so forth can also globally distort the 'lens of cognition'; for example Jim may cognise everything in terms of being watched by aliens. So how can appeal to the feature of reflexivity be what distinguishes in scale the delusion of self from clinical delusions and so forth? I would reply that such pathologies could well involve *disturbances to the sense of self*, such that relative to the norm, the degree of cognitive distortion is far greater. I am suggesting that the norm itself carries reflexive layers of distortion that can be removed, that we are dealing with a sliding scale.

56. While variations of this method are to be found in modern mindfulness techniques, detailed instructions on mindfulness and meditation are sourced in the Pāli suttas, *e.g.* in The Foundations of Mindfulness (*Satipaṭṭhāna Sutta*, MN 10, 1995,transl. Ñāṇamoli and Bodhi). I have provided here only a barest outline of the methods and processes that would go into undoing the illusion of self.

is viewed, far more accurately, as stripped of affective valence.[57] And third, unmediated by the illusion of a solidly bounded axiologically salient self, one's apprehension of their psycho-physical existence in relation to the world would be far more accurate and direct, quite possibly enabling one to cognise a level of interconnectedness with the surrounding world to which people are normally blind. It is perhaps no accident that one of the most frequent depictions of the cognitive shift that occurs in *nibbāna* (and in preliminary flashes of insight along the way) is that of apprehending the underlying interconnectedness and 'oneness' of things, with one's psycho-physical existence experienced as somehow integrated with, rather than separately salient from, the world around.

With a considerably less distorted mode of cognition, it is not unreasonable to suppose that the mind would become, in certain respects, a tool of greater precision, lending various claims a proportionally greater credence. Perhaps facts about the nature and workings of the mind, previously obscured, would become accessible. The gaining of insight knowledge would then not only be a matter of eliminating cognitive distortion around the delusion of self, but also of activating capacities previously inhibited by the delusion. Such capacities may indeed help to explain why it would be possible to function autonomously in the world without the sense of self.

Conclusion

While no Buddhist practitioner will know as much as Mary Analogue about details surrounding the illusion of self and its cognitive dissolution, I have been supposing that most practitioners will, prior to full awakening, have at least some knowledge of the proposition that there is no self. But what are we to say about cases where there is little or no such prior knowledge or indeed, belief? It is not in fact clear that there could be any such cases. The meditative practice set out by Buddhist tradition includes a rigorous combination of attentional training and reflection, which may well require endorsing the thought that there is no self. Hence, the exact role of prior intellectual knowledge or belief in effecting the cognitive transformation to *nibbāna* remains unspecified. But regardless of whether there must be prior intellectual knowledge or belief that there is no self, I contend that all cases of awakening — which by definition involves a dispelling of the cognitive illusion and delusion that one is a self — will be cases that exemplify great epistemic improvement, even if they are not to be described as improving *existing* knowledge that there is no self. Here is why.[58]

First, there will be the supplanting of a false, action-based framework belief with the correct framework belief that there is no such self (along with a usurping of the entire web of *taṇhā*-driven cognitions by those consistent with there being no self). The event of having completely seen through the delusion of self will be marked by a (veridical) feeling of profound noetic resonance. Second, I contend that as a result of this transformation, the arahant will have come to harbour a doxastically integrated belief that there is no self. The action-based component will stem from thinking, feeling, speaking and behaving in a way that is congruent with the reality of no self (including delivering practical teachings commensurate with the truth of no-self). The judgement-based component will stem from having dispelled the delusion of self, which involves — via the feeling of noetic resonance — recognising the illusion of self, in some way, *as* an illusion.[59] Perhaps the thought that there is no such self will not be explicitly entertained, just as we may not explicitly entertain the thought that the contents of a dream we have woken up from are not real. The judge-

57. In relation to studies on the effects of meditation increasing the level of attentive awareness and decreasing emotional distortion, Davis and Thompson propose that cognitive science should further investigate the Buddhist claim that "mindfulness counteracts not knowing, by increasing awareness of presently arising stimuli, and also counteracts knowing wrongly, by attenuating emotional distortions of attention, perception and memory" (2013, 594–595).

58. Similar considerations may apply to cases where someone without prior intellectual knowledge of the relevant proposition sheds a (doxastic) clinical delusion, phobia or superstition.

59. As expressed in this conclusion and in note 33, the role that prior intellectual knowledge must play in attaining such understanding — along with the mechanics of the understanding itself — remains unclear.

ment-based component will then be to some extent implicit, just as our belief about the unreality of a dream is implicit. If asked explicitly whether they believe that they harbour a bounded, separate self, the arahant, on the basis of the experience of having overcome the delusion, will reflectively endorse the proposition that there is no such self to be harboured. Third, the experience of having overcome the delusion of self may well serve as good *justification* for their belief that there is no such self. In that case, the arahant is likely to know that there is no self.[60]

Finally, the awakened arahant will have eliminated the layers of pervasive cognitive distortion that attend the delusion of the self. Perhaps this loss of distortion will lend extra credence — and indeed, further justification — to their belief about the non-existence of self. If these speculations are correct, then all cases of gaining complete insight into the reality of no self, whether requiring prior knowledge or not, will be cases where the *epistemic status* of the individual has markedly improved. By acquiring doxastically integrated knowledge of no self, through revising one of their deepest framework beliefs, the arahant will have come to occupy, as Strawson has put it, a "more correct view of the world".[61]

60. This part of proposal is especially tentative; to reiterate note 34, the question of what sort of evidential and justificatory role might be played by a direct and veridical experience of having overcome the illusion of self (or for that matter, any doxastic delusion, phobia, or superstition, etc.) is a topic for further investigation.

61. Through the evolution of this paper a number of people have been of particular help. I first wish to thank Christian Lee, Noa Latham, and the anonymous referees for their meticulous and insightful comments on recent drafts of the paper, and John Maier for his critique of an earlier introduction. I also thank audiences at Jawaharlal Nehru University and the University of Calgary at which I presented an early version in 2010: in particular, Manidipa Sen, Phil Gerrans, David Chalmers, David Dick, Christopher Framarin, and John Baker. Galen Strawson was very helpful in discussing the core ideas at their inception, as was Rebecca Collins, Nic Damnjanovic, and Jane McKessar. I am finally grateful to my mother, Sonia Albahari, for the many conversations that throughout the whole process helped me to improve the paper.

References

Albahari, Miri. (2006), *Analytical Buddhism: The Two-Tiered Illusion of Self*. (Basingstoke: Palgrave MacMillan).

—. (2011), "Nirvana and Ownerless Consciousness," in Mark Siderits, Evan Thompson and Dan Zahavi (eds) *Self, No Self? Perspectives from Analytical, Phenomenological and Indian Traditions,* (Oxford: Oxford University Press).

—. (2014), "Alief or Belief? A Contextual Approach to Belief Ascription," *Philosophical Studies*, Vol 163: 3, 701–720.

Bodhi, Bhikkhu. (1994), *The Noble Eightfold Path: The Way to the End of Suffering* (Kandy, Sri Lanka: Buddhist Publication Society).

—. (1995), *The Connected Discourses of the Buddha: A Translation of the Saṃyutta Nikāya*. (Boston: Wisdom Publications).

Bolte Taylor, Jill. (Feb 2008), "My Stroke of Insight," retrieved from http://www.ted.com/talks/jill_bolte_taylor_s_powerful_stroke_of_insight.

Bortolloti, Lisa. (2010), *Delusions and Other Irrational Beliefs*. (New York: Oxford University Press).

Damasio, A. (1999), *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. (New York: Harcourt Brace).

—. (2012), *Self Comes to Mind: Constructing the Conscious Brain*. (New York: Pantheon Books).

Davidson, D. (1984), *Inquiries into Truth and Interpretation*. (Oxford: Clarendon).

Davis, Jake, and Thompson, Evan. (2013), "From the Five Aggregates to Phenomenal Consciousness: Toward a Cross-Cultural Cognitive Science," *Blackwell Companion to Buddhist Philosophy*.

Davidson, Richard J and Lutz, Antoine. (Jan 2008), "Buddha's Brain: Neuroplasticity and Meditation." *IEEE Signal Processing Magazine*, 171–174.

Dennett, D.C. (1991), *Consciousness Explained* (London: Penguin Books).

—. (1987), *The Intentional Stance* (Cambridge, MA: MIT Press).

Double, Richard. (1991), "Determinism and the Experience of Freedom," *Pacific Philosophical Quarterly* 72, 1–8.

Farb, N.A; Anderson, A.K; Mayberg, H; Bean, J; McKeon, D; Segal, Z.V. (2010), "Minding One's Emotions: Mindfulness Training Alters the Neural Expression of Sadness." *Emotion*, Vol 10, No 1, 25–33.

Flanagan, Owen. (2011), *The Bodhisattva's Brain: Buddhism Naturalised* (Cambridge, MA: The MIT Press).

—. (1992), *Consciousness Reconsidered* (Cambridge, MA: The MIT Press).

Frankish, Keith. (2009), "Delusions: A Two-Level Framework." In M. Broome and L. Bortolotti (eds.), *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*, (Oxford: Oxford University Press) pp. 269–284.

Gendler, Tamar Szabó (2008a), "Alief and Belief," *Journal of Philosophy* 105, pp. 634–63.

—. (2008b), "Alief in Action (and Reaction)," *Mind & Language* 23, pp. 552–85.

Gertler, Brie (2011), "Self-Knowledge and the Transparency of Belief" in Anthony Hatzimoysis (ed) *Self-Knowledge*, (Oxford: Oxford University Press).

Hetherington, Stephen. (2006), "Scepticism and Ordinary Epistemic Practice," *Philosophia* 34, 303–310.

—. (2001), *Good Knowledge, Bad Knowledge: On Two Dogmas of Epistemology*. (Oxford: Oxford 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).

Jackson, Frank. (1986), "What Mary Didn't Know." *Journal of Philosophy* (83): 291–295.

James, W. (1890/1981), *The Principles of Psychology: Vol. 1*. (Cambridge, MA: Harvard University Press).

Kitchener, Richard F. (2003), "Review of Stephen Hetherington 'Good Knowledge, Bad Knowledge: On Two Dogmas of Epistemology.'" *Philosophical Books*, Vol 44, Issue 4. 363–365.

Kriegel, Uriah. (2012), "Moral Motivation, Moral Phenomenology, and the Alief/Belief Distinction," *Australasian Journal of Philosophy*, Vol 90, No.3, 469–486.

Lebrecht, Sophie; Bar, Moshe; Barrett, Lisa F; Tarr, Michael J. (2012), "Micro-valences: Perceiving Affect Valence in Everyday Objects." *Frontiers in Psychology*, Vol 3: 107.

Leite, Adam. "Epistemic Gradualism and Ordinary Epistemic Practice: Response to Hetherington." *Philosophia*, 34, 311–324.

Metzinger, T. (2003), *Being No One* (Cambridge, MA: MIT Press).

Nida-Rümelin, Martine. (2010), "Qualia: The Knowledge Argument," *The Stanford Encyclopedia of Philosophy* (Summer 2010 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2010/entries/qualia-knowledge/>.

—. (2004), "What Mary Couldn't Know: Belief About Phenomenal States" in Peter Ludlow, Yujin Nagasawa and Daniel Stoljar (eds) *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*. (Cambridge: MIT Press).

Ñāṇamoli, Bhikkhu, and Bodhi, Bhikkhu. (1995), *The Middle Length Discourses of the Buddha: A Translation of the Majjhima Nikāya*. (Boston: Wisdom Publications).

Panksepp, Jaak. (1998), *Affective Neuroscience: The Foundations of Human and Animal Emotions*. (New York: Oxford University Press).

Parfit, Derek. (1984), *Reasons and Persons*. (Oxford: Oxford University Press).

Ryle, Gilbert. (1966/1994), "Self-Knowledge" in Quassim Cassam (ed) *Self-Knowledge*, (Oxford: Oxford University Press).

Schwitzgebel, Eric (2010), "Acting Contrary to Our Professed Beliefs or The Gulf between Occurrent Judgment and Dispositional Belief." *Pacific Philosophical Quarterly* 91, 531–553.

Smith, M. (1994) *The Moral Problem* (Oxford: Blackwell).

Sommers, Fred. (2009), "Dissonant Beliefs," *Analysis* 69, 267–74.

Stalnaker , Robert. (1984), *Inquiry*. (Cambridge, MA : MIT Press).

Strawson, Galen. (2010), *Freedom and Belief*, (Oxford: Oxford University Press, 2nd ed).

Van Inwagen, Peter. (1983), *An Essay on Freewill*, (Oxford: Oxford University Press).

Velleman, J. David. (2002), "Identification and Identity," in Sarah Buss (ed.), *Contours of Agency: Essays on Themes from Harry Frankfurt*, (Cambridge MA: MIT Press).

—. (2000). "The Aim of Belief" In *The Possibility of Practical Reason*, (New York: Oxford University Press).

Williamson, T. (2000), *Knowledge and Its Limits,* (Oxford: OUP).

Wittgenstein, Ludwig. (1969), *On Certainty,* (Oxford: Basil Blackwell).

Zahavi, Dan. (2011), "The Experiential Self: Objections and Clarifications," in Mark Siderits, Evan Thompson and Dan Zahavi (eds) *Self, No Self? Perspectives from Analytical, Phenomenological and Indian Traditions,* (Oxford: Oxford University Press).

Zimmerman, Aaron. (2007), "The Nature of Belief," *Journal of Consciousness Studies* 14.11, pp. 61–82.