CONSTRUCTIVE-ENGAGEMENT DIALOGUE
PHILOSOPHY OF MIND

AUTHOR MEETS CRITICS:

**THE ILLUSION OF SELF REVISTED:
REPLIES TO CRITICS**

KARSTEN J. STRUHL

ABSTRACT: *Anand Vaidya, Sean Smith, and Mark Siderits have presented thoughtful comments and provocative challenges to my article “What Kind of an Illusion is the Illusion of Self?” Their challenges raise significant questions about the nature of illusion, whether Buddhism is denying the self in all senses of the term, whether there could be a self that exists for some limited duration of time and has at least some measure of control, whether there is a phenomenal illusion of self, whether the neuropsychological assumptions embedded in Thomas Metzinger’s Phenomenal Self Model is consistent with Buddhist metaphysics, the usefulness of evolutionary psychology in explaining why we have the illusion of self, whether the I-sense is a result of natural selection or cultural selection, vipassanā meditation as a form of verification and its usefulness for extinguishing the I-sense. The discussion here is my response to these criticisms through which I further clarify and develop my arguments and, in some ways, amend my position*

Keywords: *Anand Vaidya, Buddhism, evolutionary psychology, I-sens, illusion of self, Mark Siderits, neuropsychology, phenomenal self model, physicalism, Sean Smith, vipassanā (mindfulness meditation)*

I am honored to have Anand Vaidya, Sean Smith, and Mark Siderits comment on my article. Since their comments and criticisms overlap, I will be organizing my responses to them by grouping their criticisms within a number of topics, often responding to more than one of them within a given topic. In assessing their responses to each of my claims, I intend to do the following: (a) I will seek to clarify my claims when I think they are misunderstood; (b) I will attempt to respond to arguments that I think require

STRUHL, KARSTEN J.: Lecturer, New School for Public Engagement, USA. Email: kastruhl@tiac.net

more argumentation on my part; and (c) I will consider ways in which I might need to revise my position in light of their arguments.

1. WHAT IDEA OF THE SELF IS BEING DENIED AND THE *PĀLI* CANON

I begin with a discussion of what is being denied in Buddhism, and I reaffirm that while the idea of the self that fits the nine criteria elaborated in my article is being denied, the self is not being denied in all senses of the term. Smith says that my using the *Pāli* canon as a starting point for this is suspect, since it is a mistake to claim “that the *Pāli* suttas give us any clear understanding of what the historical Buddha actually thought.” But I am making no such claim. I do not claim to know whether these suttas accurately reported what the Buddha said in general or specifically about the self or even if there was a historical Buddha. Furthermore, even if these suttas accurately recorded what the historical Buddha said, they would still not tell us what he actually thought. But all this is not relevant to my starting point. The reason that I begin with the *Pāli* canon is because all Buddhist schools recognize its suttas as the words of the Buddha and, therefore, the criteria for the self that is being denied by these suttas would be accepted by all of them. Vaidya, however, claims that there are significant variations across schools, suggesting that they vary on “whether the self is momentary or that there is no subject of experience.” I have two replies to this. First, it is the *dhammas* and not the self which are momentary, and all Buddhist philosophical schools accept the momentariness of the *dhammas* (however they may differ on the ontological status of the *dhammas*). Second, while there is some disagreement about how to characterize the subject of experience, none of the philosophical schools regard the subject of experience as a self, as an entity which observes my experiences and yet is distinct from them. In short, while demonstrating this by examining the way in which each of the schools deny the self is beyond the scope of this response, I believe that all Buddhist philosophical schools, however they might differ in particulars, would deny that there exists a self that has any of the nine characteristics that I indicate.¹

Smith raises a question about my claim that I am not denying the existence of a self in all senses of the term. He equates this with my saying that “the Buddha wasn’t a full-blown anti-realist about the self” and that I am “portraying the suttas as agnostic about the existence of the self.” It is important for me to make clear that I am not claiming that Buddha was agnostic about the existence of an entity which has the nine characteristics that I discuss at the beginning of my article. Buddha, as portrayed in the *Pāli* Canon, was, I believe, indeed a full-blown anti-realist about there being any such entity. However, it is often the case that when people first learn of the *anattā* doctrine, they are perplexed, thinking that what is being denied is the non-existence of themselves as persons. It then needs to be explained that there are indeed persons

¹ I will, however, concede that one of the approximately 18 schools of early Buddhism – the Pudgalavādin school – did have a concept of the person which imports at least one of the characteristics of the self which are denied. We know about them primarily through the critique of their position by Vasubandhu. However, this school did not survive, and the texts that we have which represent the thinking of early Buddhism is the *Pāli* canon of the Theravada school or the Chinese agamas.

understood as a set of aggregates which have both synchronic and diachronic continuity. At this point someone may declare that such a set of aggregates is a self, specifically a process self. Indeed, some philosophers of Buddhism have suggested that Buddha does not deny the existence of a process-self.² The process “self” is the sum of the psychophysical processes that constitute the person, and so, as I say in my article, in order to avoid confusion it is better to use the term ‘person’ for the sum of these processes; and the person, while conventionally real, is not ontologically real. My overall conclusion is that Buddha is indeed an anti-realist as regards the existence of the self as an ontological entity, but he did not deny the existence of the self in all senses of the term.

Vaidya brings up another worry as regards the nine characteristics that I claim are being denied by Buddhism. “...the reasons that go into showing that one kind of self is an illusion don’t automatically transfer over to saying that another kind of self is an illusion. For example, just because there isn’t a permanent self across time, it won’t follow that there isn’t an independent self for some portion of time.” And conversely, that “the self isn’t independent, because everything is dependently originated, isn’t sufficient for showing that there is no controller self or self as subject of experience that last longer than a moment.” This may, in fact be correct, but I do not think that it matters. What I have tried to do is draw out nine characteristics of self that I take to be denied by all the schools of Buddhism and to whose denial at least most Buddhist thinkers would agree. Furthermore, while it is possible to analytically separate these characteristics, many of them do mutually implicate each other. For example, an ontologically independent self would need to be a unitary and bounded self. Also, since the boundary lines of the self designate the body and its experiences, it would be a self that owned these things.

2. THE ARGUMENTS

In my article, I mentioned what I take to be the strategy of the main philosophical arguments from the *Pāli* Canon for the Buddhist denial of self. I characterized these arguments as a reductionist strategy which reduces the person to a set of psychophysical process – the *skandhas*, or aggregates. I point out that the reductionist strategy develops two parallel arguments which I call *criterion* arguments, arguing that the aggregates both individually and collectively cannot fulfill the criteria of what it means to be a self. The two criteria are, respectively, the criterion of permanence and the criterion of control. Vaidya challenges each of these arguments. On the criterion of permanence, Vaidya grants that the aggregates are constantly changing, but he argues that this does not rule out the possibility that there could still be a semi-permanent self, a self that “is a phase of a person. Something that exists for some amount of time, but not for all time, and is a subject to change as well, but serves other cognitive roles during each phase, such as decision making and being the subject of experience, accounting for diachronic

² See, for example, Gowans 2003, 91-103. Gowans argues that while the Buddha did deny that there were substance-selves, he believed that we are process-selves until we attain *Nirvāṇa*.

knowledge...” His argument for this employs Husserl’s phenomenology of internal time consciousness to insist that our time sense includes pro-tension (intention toward the future) and re-tention (intention toward the past) and that, therefore, there must be duration beyond the moment. Thus, he concludes that the Buddhist argument which employs the criterion of permanence does not prove “that there is no sense of self at all that has some duration of time, longer than a moment, and less than forever.” On the argument based on the criterion of control, Vaidya argues that we experience at least limited control without which we could never have formed the concept of control which is being denied; and such control, he argues must presuppose a self, an agent who has at least some measure of control.

I will take the argument for the agent who exercises control first. It is certainly true that we experience intentions, decision making, and actions to fulfill these intentions. But to conclude from these experiences that there must be an agent who intends, decides, and acts begs the question. What we can properly conclude from these experiences is that there are intentions, decisions, and actions; and the Buddhist understanding of dependent origination explains how these come about as a result of previous psychophysical processes. The idea that they require an agent that is in some sense separate from the stream of these processes presupposes the existence of such an agent; but that is precisely what is in question. The problem in modern western philosophy goes back to Descartes who argued that since I can doubt and since doubt is a form of thinking, there must be an “I” who is thinking. But what Descartes should have concluded is simply that there is thinking going on. Similarly, what we can conclude from our experience of intending and deciding is that these experiences are occurring as part of the ever changing stream of experience. The feeling of control comes not from our experience of a self-determining entity, but our experience of some of the aggregates exercising control over other of the aggregates – what Siderits, who Smith refers to in his reflections on my article, calls “a shifting coalition” of physical and psychological process in which at any given time some of these processes performs the executive function in relation to other parts, but there is no part which is always the controller. What I suspect is behind Vaidya’s challenge is the conflation of self with what Buddhism calls ‘person’. However, the person is only conventionally real, because it is constructed from psychophysical processes – momentary *dharma*s – which are, at least for early Buddhism, ultimately real. This echoes the idea of the person proposed by the Pudgalavādin school – one of the schools of early Buddhism which did not survive – and which is now rejected by every existing school of Buddhism.³ I think that the same conflation of person with self is behind Vaidya’s proposal of a “self” that is a semi-permanent “phase of a person” whose aggregates are constantly changing. Why not simply say that the person is the totality of the changing aggregates throughout one’s life.

Let me now say something about being the subject of experience and Vaidya’s argument which employs Husserl’s phenomenology of internal time consciousness. I certainly experience myself as the subject of my experience in the sense that my

³ This is not to say that all contemporary Buddhist thinkers reject this idea.

experience is from a certain perspective. But that this perspective is inhabited by a self is what is in question. When I look at my computer screen at the moment, I see it from a certain spatial perspective. In general, I would say that it is from the perspective of my body, but that it is *my* body is a convention of language, not a metaphysical truth.

The question of what is the best analysis of internal time consciousness is more complex, and I admit having some sympathy for Husserl's account. Husserl's phenomenology attempts to analyze the lived experience of the continuity of time as opposed to the idea of a quantitative objective metaphysical account of time. Once we perform the phenomenological reduction, we are aware of experiencing a succession of mental states in which our sense of how fast or slow time moves does not coincide with clock time. This is reflected in ordinary speech when we say that the time moved very slowly while listening to a boring lecture. What we experience is not an objective succession of "nows" in objective time but "now" as a point of orientation from which one experiences perceptions, feelings, memories, expectations, fantasies, hopes, etc. as belonging to the present, past or future. Thus, our sense of "now" already "re-tends" a sense of what is past and "pro-tends" what is to come. Listening to music provides a good way of seeing this, as the sound we are hearing now is a melody only if it retains the sound of the past and anticipates the sound to come. However, the appeal to Husserl's phenomenology of internal time consciousness will not help Vaidya's proposal. What it helps us to understand is how the pro-tension and re-tension of the structure of time consciousness can still exist *in the moment* as memory and anticipation. Put another way, since consciousness is intentional for both Buddhism and Husserl, *each moment* of lived experience intends both the past and the future. Thus, the "self" that is the "phase of the person" or perhaps the whole of the person through life is a causally connected succession of a momentary constellation of psychophysical processes, each of which manifests an intentional structure toward the past and the future. There is no self which has the characteristics that are being denied by Buddhism which inhabits the person. The "person" identified as the totality of these processes is conventionally real, but since it is composed of and constructed by the *dhammas*, is not ultimately real.⁴

Before I leave the discussion of the arguments for no-self, I want to respond briefly to a criticism that Smith makes concerning these arguments. Smith notes my claim that the criterion arguments against the self are not sufficient unless we add the exhaustiveness claim, the claim that there is nothing more to the person than the five aggregates. However, he takes issue with my entertaining the possibility that the self could be something beyond the aggregates and my claim that we need to argue against this. In contrast, Smith insists that the exhaustiveness claim is already provided in the Pāli suttas, and he quotes one of the suttas which says that the brahmins who are being challenged by Buddha's arguments regard the self as either one of the aggregates or all of them collectively. I take Smith's point to be that we need not entertain the idea of a

⁴ Whether or not the *dhammas* are ultimately real is open to debate. While all schools of Buddhism would agree that the person is empty of self and cannot be ultimately real, the *dhammas* are ultimately real for the Abhidharma of the Pāli canon but not for Madhyamakas for whom the *dhammas* are also empty of intrinsic nature.

self beyond the aggregates, since those who claim that there is a self do not identify it with anything beyond the aggregates. If this is what Smith has in mind, I find his position puzzling for two reasons. First, as I pointed out in my article, the Buddha was surely aware that traditional Brahmanical thinking posited the *ātman* as the true self which transcends the individual person and which was both permanent and the ultimate controller. Secondly, even if Buddha or his disciples were not attempting to challenge the idea of a self that transcended the aggregates, it is certainly reasonable to consider this today. This is not a question of being a Buddhist modernist or of embracing modern scientific anti-realism about the self but of making the best philosophical arguments for the claim that there is no self.⁵ However, my point in developing these arguments was not to definitively prove by philosophical argument alone that there is no self, and my discussion of these arguments is essentially an attempt to retell briefly what I take to be some of the main arguments that have been made by others. As I make clear in the article, I do not think that the doctrine of no-self can be established by argument alone. My purpose in recapitulating these arguments in my article is not to defend any particular Buddhist version of them but to make plausible and lay the groundwork for my discussion of the self as an illusion.

3. WHAT IS AN ILLUSION?

In my article I defined illusion in three ways – as involving motivation (beliefs based primarily on wishes and desires), as a cognitive illusion (a false belief which is so interwoven with other aspects of our cognitive system that it is difficult to dislodge), and as a phenomenal illusion – and I claim that the self is an illusion in all three senses. Since Vaidya defines illusion as something perceptual and argues that the self is not something that we can perceive or misperceive, it cannot be an illusion in any of the three senses. This seems to be an attempt to win the argument by definition. While it is true that it is common in psychology to focus on perceptual illusions, there are many other senses of the term often used by philosophers, by psychologists, by social and political theorists, and by people in ordinary speech. I have already in my article cited Freud’s use of the term for my first definition of illusion as motivated by human wishes. Here is another example of illusion based on wish from Marx. “The abolition of religion as people’s *illusory* happiness is the demand for their *real* happiness” (Marx 1994, 28). Here is an example of its use in contemporary psychology that is not about perceptual illusions. “The illusion of control is the tendency for people to overestimate their ability to control events; for example, it occurs when someone feels a sense of control over outcomes that they demonstrably do not influence” (Wikipedia). Here is a current example, quoted in the Merriam Webster Dictionary, of the political use of the term in daily newspapers during the year 2020. “All week long, Republicans at the nonconvention convention tried to create the *illusion* that the pandemic is largely a

⁵ In modern western philosophy, raising this question comes easily after Descartes’ cogito and Kant and Husserl’s transcendental ego. In modern Hindu philosophy, the argument for no-self needs to confront the philosophy of Advaita Vedānta. I am not, after all, addressing ancient Buddhist thinkers but modern readers who may be skeptical of the Buddhist claim that there is no self.

thing of the past”. I would add to these uses, illusions about personality traits and character dispositions. I may think of myself as a kind, generous, and cooperative person, but, if this is not true, it is reasonable to say that I am under the illusion that I am kind, generous, and cooperative. In short, Vaidya’s definition of illusion as applying only to perceptual misrecognition seems unnecessarily and radically restrictive. In addition, the substantive question cannot be settled by definition. Even if we decide to restrict the term ‘illusion’ to perceptual illusions, we are still left with questions about the ways in which reality is misrepresented because of our desires and wishes, because of our cognitive apparatus, because of our language structure, or because of the way things appear to be. If we do not want to apply the word ‘illusion’ to such cases, then we can simply use another word, for example ‘illuziya’ which is the transliteration of the Russian word for illusion (‘иллюзия’). Having made that distinction, we could still ask the questions – Is the self an illuziya? If so, what kind of an illuziya is it?

Vaidya, of course, insists that the self is not an illusion (or even an illuziya) in any of the three senses which are the focus of my discussion, since he makes clear that he thinks that there is a self which, while not permanent, is at least semi-permanent and exercises many of the cognitive functions that are traditionally ascribed to it. I have already given my response to this in the previous section, so I will not discuss again here. But I do want to say something about what Vaidya considers a paradigm case of a perceptual illusion – mistaking a coiled rope for a snake. While this is a clear case of something appearing as other than what it is, it is not a perceptual illusion simpliciter. As Vaidya admits, I see the coiled rope as a snake not merely because it looks somewhat similar to a snake but because “my *fear* [italics mine] of the latter triggers a foul up in my perceptual processing;” and fear is motivated by the desire or *wish* to be safe. It is the wish to be safe and the fear of snakes that *motivates* my seeing the coiled rope as a snake. Furthermore, there would not be a fear of snakes unless one had a concept which groups a variety of animals into one classification – snake – and has learned through experience to fear them. In all, this “perceptual” misrecognition has built into it both motivational and cognitive elements.

Nonetheless, Vaidya raises what might seem to be a more substantive issue using the rope-snake illusion. By coming closer, I can recognize that it is a rope and, thus, extinguish the illusion. This, however, requires that there be someone for whom the illusion is extinguished, someone for whom “extinguishing the illusion is a case of diachronic knowledge change.” But that someone is *me*. This, Vaidya asserts, entails that there must be “a self that had an epistemic change in light of a perceptual change.” If there is no self, for whom is the illusion extinguished? To answer this question, I will repeat what I have said in my article. Diachronic unity – and hence, diachronic knowledge – can be explained by the concept of dependent origination, which means that it is not necessary to resort to the concept of self to explain it. Thus, to answer the question for whom is the illusion extinguished, we need again to distinguish the concept of a person from the concept of a self. It is the *person* who comes to recognize that what appeared as a snake is really a rope. However, the person is a conventional reality, and the indexicals ‘I’ or ‘me’ are simply useful ways of speaking. In short, what is

really happening is a set of continuous impersonal processes connected causally through time. Here is one (admittedly) linguistically awkward rendition of what is really happening. Certain psychophysical processes occur at time x such that certain fears and beliefs in conjunction with distance from the coiled rope generate a perceptual snake-like appearance. At time y (later than time x), as the physical organism approaches the rope, new perceptions occur which override the fear and create beliefs that what is seen is not a snake but a rope. On this rendition, the answer to the question for whom is the illusion extinguished is: “for no one.” The illusion of the snake is extinguished, but there is no one for whom it is extinguished. There was perceiving and thinking that the rope was a snake, and now there is perceiving and thinking that the rope is not a snake; but there was and is no perceiver or thinker.

4. THE SELF AS A PHENOMENAL ILLUSION

I think what is really under attack in Vaidya’s analysis of illusion is my claim that the self is not only a motivational and cognitive illusion but a phenomenal illusion. Vaidya suggests that my model for extinguishing the illusion of self is the Kanizsa square illusion, and offers two criticisms of this. First, one can train oneself to dispel the Kanizsa illusion (Vaidya switches from the Kanizsa square illusion to the Kanizsa triangle illusion) by seeing the black objects as packmen or as slices of pizza. This can be done, he claims, by uttering the mantra “packman, packman” as we look at the figure. If the mantra works, this would make the Kanizsa illusion different than the Müller-Lyer illusion which is perceptually impenetrable. Second, referring to Hume’s introspective analysis, Vaidya argues that since we do not perceive the self in the first place, there is no way to misperceive it, and, hence there is no way to correct the perceptual illusion. Thus, in contrast to the rope-snake or the Kanizsa square illusion, both of which can be corrected, the self cannot be a perceptual illusion; in which case, since Vaidya claims phenomenal illusions are perceptual illusions, the self cannot be a phenomenal illusion. I shall consider this second criticism first.

In my article, I explained a phenomenal illusion as something which appears differently than what it is, and, when introducing the term, I did say that this could also be called a “perceptual” illusion. Given Vaidya’s second criticism, I must now amend this as follows. A phenomenal illusion is an illusion in which what exists appears to be something other than what it is, but, while a perceptual illusion is a phenomenal illusion, not all phenomenal illusions are perceptual illusions. The word ‘appears’ can have two meanings. In the first case, it is indeed perceptual, like the Kanizsa square or the rope-snake illusion. However, we can also use ‘appears’ in the sense that some experience or set of experiences appears to the mind to be something other than it is, because of the way the mind spontaneously structures or reacts to that experience. Thus, certain mental phenomena may create the feeling or intuition that they reveal something that does not exist. In some cases, this feeling or intuition may be based on sensory data which is not processed consciously; for example, I might have feeling that there is danger in this environment without knowing why I have this feeling (and there might be, in fact, signs of danger which I was only aware of subliminally). In

other cases, it may be based on certain specific emotions. I can feel that I am in love, but it may be really the desire to be loved that I feel. I may experience an emotion that I think is anger, but careful attention to the emotion may reveal that it is really fear. I may have the feeling that I am in control, but it is an illusion based on my failure to recognize that I have been manipulated. And so, in the case of the self, I have an immediate feeling or intuition that there resides deep inside me something unitary and enduring which is also a control center and owner of my experiences and which demarcates me from others. This is the phenomenal illusion of self which, although pre-conceptual, tends to produce a belief that the self exists. Siderits defines it nicely in his response to my article. “To call the self a phenomenal illusion is to say that our experience possesses the phenomenal character that leads to our judging that we have a self.” However, for reasons that I will soon discuss, Siderits denies that there is such a phenomenal illusion.

To return to Vaidya’s second criticism, it is certainly true that when we introspect we find that there is no *perception* of the self to be found. Thus, if the self does not exist, all we have is a mistaken belief which our introspective experience can disconfirm. That this analysis leads to the conclusion that the self cannot be a phenomenal illusion is odd, since Buddhism also recognizes that in *vipassanā* meditation, which is a much more controlled form of introspection than anything Hume had in mind, there is no self to be found. However, there is something observed in *vipassanā* meditation which disconfirms the I-sense – the arising and disappearing of various mental and physical processes which either individually or collectively may be mistaken for the self.⁶ The phenomenal illusion of self is the spontaneous feeling that these processes are “mine” or that “I” am in control of them and that there is some essential “me” that does not change even as these processes continue to arise and disappear; and this leads easily, if not also spontaneously, to a set of metaphysical beliefs that misrepresents the person as having a self.

Let me now turn to Vaidya’s first criticism of the self as a phenomenal illusion. Vaidya takes me to be saying that my model for extinguishing the illusion of self is the Kanizsa square illusion which by careful attention to the space between any of the two black objects can be seen to be an illusion but which then reappears as soon as one steps back from that careful attention and looks at the figure as a whole. Here again is the Kanizsa square illusion.

⁶ One other difference between Hume and Buddha is that, for Buddhism, these psychophysical processes are causally connected.

I do indeed use the Kanizsa square illusion to discuss what it might mean to have a phenomenal illusion and the difficulty of extinguishing the illusion of self for very long at the phenomenal level. However, I need to clarify how I understand the relation between the phenomenal illusion of self and the Kanizsa square. I use the Kanizsa square illusion *not as a model* of the phenomenal sense of self *but as an analogy* to explain both how an illusion can continue to exist even when we know that it is an illusion and the difficulty of maintaining a consistent non-self perspective at the phenomenal level, to explain why even after we move further along path, we would still oscillate between a non-self and a self-perspective. The point of this analogy is to say that just as we can oscillate between seeing that the Kanizsa square is an illusion through careful attention to the space between the black objects and then having the illusion reappear when we look at the figure as a whole, so through the careful attention in *vipassanā* meditation, we can come to extinguish the phenomenal illusion of self only to have it reappear when we step back from the attentional focus. If the Kanizsa square illusion would be a model rather than simply an analogy for the illusion of self, then the phenomenal illusion of self would be a perceptual illusion; but for the reasons that I have discussed above, it is not a perceptual illusion.

Look again, dear readers, at the whole figure which constitutes the Kanizsa square illusion. Does the illusion of the square really disappear when one says “packman, packman” repeatedly? Vaidya claims that the mantra works to dispel the illusion. I have tried this mantra and found that it does not work, so at this point I can only invite my readers to do the experiment. And if it does seem to work at least momentarily, let me ask my readers if they are really observing the figure as a whole or, through the hypnotic effect of the mantra, are instead shifting the focus from the figure as a whole to the space between two of the black objects; or perhaps shifting so rapidly from a focus on one space between the black objects to another that it is easy to fool oneself into thinking that one is observing the figure as a whole.⁷ However, suppose that Vaidya is correct in claiming that it is possible to penetrate the illusion even when stepping back and looking at the figure as a whole. What would follow from this is simply that it is a bad analogy, not that my analysis of the phenomenal illusion of self is wrong. I would simply need to choose another example for the purposes of analogy – perhaps the Müller-Lyer illusion which Vaidya grants is not penetrable even when one knows that the lines are equal.⁸

I mentioned earlier in this section that Siderits, whose definition of a phenomenal illusion of self I quote approvingly above, nonetheless does not think that the illusion of self exists at the phenomenal level. So if there is no illusion of self at the phenomenal

⁷ Perhaps this is why Vaidya prefers the Kanizsa triangle illusion to the Kanizsa square illusion, since in the triangle illusion there are only three “packmans” and spaces between them to contend with instead of four, making it easier to fool oneself while rapidly shifting from one portion of the figure to another.

⁸ However, the Müller-Lyer illusion would present a problem as an analogy to the way *vipassanā* could extinguish the illusion, as to make the analogy work, we would need to make the meditation process analogous to using a ruler to measure the lines. Since *vipassanā* does not require anything other than one’s own mind, the Kanizsa square illusion works better as an analogy for my purpose.

level, there is no experience that “possesses the phenomenal character that leads to our judging that we have a self.” To take another one of Siderits characterizations of the self as a phenomenal illusion, there is nothing in “the felt character of our experience” that “leads us to spontaneously affirm that they are given to or for a particular experiencing subject, *me*.” Siderits argument is nuanced, especially as he does not deny that there is an I-sense. However, he insists that this I-sense is not a “delusive phenomenal character” but is instead “the name of a bundle of cognitive and affective habits that continuously reinscribe belief in a self that serves as center of meaning and value.” Thus, Siderits would agree that the I-sense tends to lead us to develop certain beliefs, including the belief that there is some entity within us that can be characterized as a self, but these beliefs are formed not by what appears phenomenally but by a set of cognitive and affective habits.

Before analyzing Siderits reasoning for this conclusion, I want to make an observation about my understanding of the I-sense that I am equating with the phenomenal illusion of self. In my article, I explained the I-sense in terms of experiential data open to phenomenological analysis. In my article, I also gave several examples of the I-sense – looking in the mirror and immediately apprehending a sense of self inhabiting the body, feeling myself as a self-determining agent, episodic memory in which when I remember a past event I remember myself in that past event, and experiencing myself as distinct from others with whom I interact. I would not deny that these experiences are bound up with certain cognitive and affective habits. However, I would claim that in addition to these habits there is also an experiential phenomenal character to the I-sense. Put simply, I think there is an immediate phenomenal “feel” to the sense of self and that we can say a number of things about this through phenomenological investigation.

I now want to examine Siderits’ arguments for his claim that the I-sense is not a phenomenal illusion of self. I would break his argument into the following components:

1. It is unlikely that there is a single heritable feature of the brain that produces the I-sense, given the many tasks that it is meant to perform; and it cannot be a straightforward result of natural selection, because it does not appear until later in childhood when the child attributes mental states to herself; and it is most likely in its full development the product of socialization to one’s culture. Thus, if a phenomenal illusion is a phenomenal presentation brought about by the way the mind naturally structures our experience, it is unlikely that there is a phenomenal illusion of self
2. When we investigate the phenomenal character of our experience, we not only do not find a self (as Hume noted) but at the most find that in addition to the intentional object of our experience, there is also the awareness of that object. However, any attempt to examine the content of that awareness is always elusive (as Brentano noted).
3. Episodic memory need not contain a phenomenal sense of self. For example, that a visual experience triggers a memory of the object as pleasurable leading to a present desire for that object can be expressed in purely impersonal terms. In other words, we can describe the process as a causal series of mental

events associated with this brain and body without employing the indexical 'I'. What leads us to think that 'I' stands for some single enduring substance is simply our habitual use of that indexical rather than something in our phenomenal experience.

4. While Siderits agrees that extinguishing the illusion of self requires more than philosophical argumentation, this is not because it requires extinguishing the phenomenal sense of self, as there is, on his analysis, no phenomenal sense of self to be extinguished. What maintains and reinforces explicit beliefs in the self is a bundle of cognitive and affective habits which constitute the I-sense. Thus, to extinguish the I-sense requires extinguishing these cognitive and affective habits.

Since I intend to discuss the first argument in the next section of this response, where I will focus on the criticisms to my use of neuroscience and evolutionary psychology, and the last argument in the concluding section, where I will focus on concerns about my analysis of extinguishing the illusion of self, I will now in the rest of this section focus on Siderits' second and third argument.

Siderits' second argument is a phenomenological one. The question then is what is it exactly that we find when we do the phenomenological analysis? I would certainly agree that we do not find a self as the object of our awareness. But what about the awareness itself? It is true that when we attempt to focus on the awareness, it eludes us. This is because to focus on it is to try to turn it into an object, since awareness is awareness *of* the object. For example, I can focus on my thinking about the problem of the self, but then my thinking about the problem of the self is not my primary awareness, because now I am thinking about my thinking about the self. So I grant that I cannot find the illusion of self through this procedure, but I do not think I need to do so. The illusion of self at the phenomenal level is not a "seeing" something that I mistake for a self, the way I might mistake the coiled rope for a snake. Nor is it an appearance of something that is not there, like the appearance of the square in the Kanizsa illusion. Rather, as I have tried to explain above, it is a strong feeling or intuition that there is something that is operating in the background that I do not directly see.

In my article, I explained the phenomenal sense of self through analogy with the perceptual illusion of the Kanizsa square, and this lends itself to the confusion that I have discussed above. I hope now that I have put this confusion to rest. Let me offer a different kind of analogy. Imagine that the I-sense is a way of organizing our experience in something like the way a computer program organizes data (again, this is an analogy only). This organizing program groups experiences into forms of self-cherishing and self-grasping. The former distinguishes my interests from the interests of others, and gives my interests a central priority. The latter entails that I take the primary object to which I am attached as *me* and must defend it at all costs.⁹ As a result, I develop a

⁹ I am indebted to Jessica Locke for pointing out the implications of these concepts. "Self-cherishing...establishes a polarized phenomenal field, structuring our world such that whatever we experience comes to us through a prism of self-interest" (Locke 2018, 252). "The self-grasping orientation...causes us to defend the interest of this metaphysically dubious and ethically problematic

variety of cognitive and affective habits that make the program of self-cherishing and self-grasping seem all the more natural. This is why we cannot “see” the self but nonetheless feel or intuit that the self exists. The phenomenal “feel” of the self is not a perceptual experience and, in fact, is not any one specific experience but permeates all my experiences, as it is the way, given the program, that those experiences are organized. It is also why it is difficult to see through the illusion, and if the mind were only this program, it would be impossible to dispel it. Thus, while I feel or intuit that there is something operating in the background, I am not entirely wrong. What is operating in the background is the program of self-cherishing and self-grasping, but, given the way this program organizes my experience, what I feel or intuit is a unitary and enduring and controlling self who is the owner of my experiences.

Let us now look at Siderits’ third argument. It is certainly true that I can describe episodic memory impersonally, without reference to any first person indexical. However, that I can describe it that way does not mean that this is my experience. In my ordinary (pre-enlightened) experience, when I have a memory of eating a mango that I enjoyed, I do not just remember the mango and the enjoyment but see *me* in the memory; that is, I not only remember the mango and what it tasted like but also remember that *I was there* and that it was *me* eating the mango. In contrast to this, there are pathological conditions where the individual remembers the scene but has no memory of being there. There is, for example, the case of R.B. who, as a result of a head trauma, said that he could recall past experiences but “that the recalled experience didn’t seem like they were experiences that happened to *him*” (Garfield 2018, 295). There are also psychological techniques where people suffering from PTSD are taught to recall the traumatic event from a third person perspective in order to reduce anxiety (Garfield 2018, 297-298). Thus, it is at least possible to have memories in which experiencing a past scene is without an I-sense. My point here is not to equate these experiences with the Buddhist experience of selflessness but to contrast such experiences with what I take to be the normal experience of episodic memory in which there is a phenomenal sense of self.

Before leaving this section, I want to explicitly amend my understanding of the I-sense. I said above that I would not deny that the experiences that I am characterizing as the phenomenal illusion of the self are bound up with the bundle of cognitive and affective habits that continuously reinscribe belief in a self. Given my account of the phenomenal sense of self as a kind of program which organizes our experiences into forms of self-cherishing and self-grasping, I would now suggest that these cognitive and affective habits and the experiential sense of self-cherishing and self-grasping mutually reinforce each other. Thus, I would amend my understanding of the I-sense to say that it is intersection and interpenetration of the phenomenal illusion of self *and* a set of cognitive and affective habits which together provide the ground for the belief in the self as an ontological entity.

self” (Locke 2018, 260). However, while she uses these concepts to discuss the phenomenological structures that organize our experience and orient us to our (pre-enlightened) life-world, the computer analogy is mine. I have used this computer analogy previously to discuss the problem of universal compassion in an earlier article that I wrote. See Struhl 2018.

5. NEUROPSYCHOLOGY, EVOLUTIONARY PSYCHOLOGY, AND THE ILLUSION OF SELF

In my article I draw on neuropsychology to help explain how the illusion of self is constructed and evolutionary psychology to help explain why it was constructed. Smith takes me to task for my use of neuroscience and both Smith and Siderits pose criticisms of my use of evolutionary psychology. Before examining their arguments, I want to make clear what my purpose was in using these disciplines.

Buddhism provides excellent resources for recognizing that the self is an illusion, both by providing philosophical arguments and methods of meditation which support those arguments. It provides an insightful analysis of the role of this illusion in causing *dukkha*, and it offers a path for extinguishing the I-sense and *dukkha*. It does not, however, say very much about *how* the illusion of self is constructed and even less about *why* it is constructed. As Siderits notes in his comments about my article, the main explanation that Buddhism gives for why the I-sense is constructed is that it is result of “beginningless ignorance” which perpetuates itself through a series of rebirths until one is enlightened. But we have no explanation for why there is this “beginningless ignorance.” Why would human beings construct an illusion if it only serves to produce suffering? As Jack Engler puts it, “If the sense of separate selfhood is illusory, a way of organizing experience that is not reality based and inevitably creates anxiety and conflict, why do we repeatedly and consistently construct experience around it....Buddhism not only doesn’t elaborate on how this sense of self is constructed, it doesn’t address why we construct it....” (Engler 2003, 77).

Engler tries to fill this lacuna through a Buddhist informed psychoanalytic approach.¹⁰ In my article, I chose to use the resources of neuropsychology and evolutionary psychology. I do not claim that these resources are sufficient, and I also suggested that there is a social and narrative dimension to the construction of self. I am also open to the help of other contemporary resources that could help explain how the illusion of self is constructed and what purpose it might serve. This is not a question of being a Buddhist modernist, although it is certainly the case that each culture and each age has reinvented and must continue to reinvent Buddhism in a way that speaks to its collective life-world. Thus, if there are good reasons to think that the self is an illusion which brings about suffering, then we should use what resources we have at our disposal to understand that illusion, to understand how it is constructed, what its function is, and how we might proceed to extinguish it. In the twenty first century, we have resources at our disposal that Buddha could not have known. Why not use them. Of course, we must proceed with caution when we do this, which is why I welcome the reflections of Smith and Siderits on my use of neuro- and evolutionary psychology. Now, to their arguments.

¹⁰ Engler, who is both a Buddhist practitioner and psychotherapist, argues that the illusion of self is a “*compromise formation* in the psychodynamic sense; a psychological structure that emerges to confront a ‘danger situation’ by binding anxiety and warding off unwanted and unwelcome knowledge and the aversive feelings that knowledge evokes” (Engler 2003, 78). The “danger situation” to which Engler refers is not the fear of death, which already presupposes the sense of self, but the fear of emptiness.

Smith argues that my use of Thomas Metzinger's Phenomenal Self Model (PSM) is inconsistent with Buddhism, because Metzinger is committed to physicalism. Physicalism, in this context, is the view that "phenomenal states are identical with brain states" and, therefore, that brain states cannot be the cause of mental states, because "they *are* mental states." From this, Smith thinks it follows that "what makes the self an illusion is that it is merely a neural simulation." On this basis, Smith argues that there is a fundamental difference between Metzinger's understanding of how the illusion of self is constructed and Buddhism's analysis. For Buddhist philosophers, "mental categories corresponding to what we would call 'consciousness' are explanatorily basic," which is to say that the reduction of the self is to "phenomenal kinds which are...metaphysically basic." The contrast, according to Smith, is straightforward. "For Buddhist philosophers, the reduction base contains phenomenal kinds, for Metzinger the reduction base is strictly neuro-physical."

I need to make clear at the outset that there is nothing in my analysis which commits me to physicalism or which posits the identity of mental and brain states. Smith's argument attributes such a physicalist metaphysics to me because: (a) he assumes that Metzinger is an identity theorist; (b) I use Metzinger's analysis of the PSM to explain how the phenomenal illusion of self is constructed; and (c) therefore, I too must be an identity theorist. But one can use someone's analysis without being committed to their metaphysics. I take no position on physicalism or identity theory. I suspect that the mind in some way supervenes on brain functions, but that does not mean that mind is reducible to the brain. I do not know if it makes sense to say that phenomenal kinds are metaphysically basic or neural processes are metaphysically basic or even that there is anything which is metaphysically basic.¹¹ In fact, I think that my refusal to take a position on this question is consistent with the *Cūlamālunkya* sutta (MN 63, I 426-32) where Mālunkyāputta, one of Buddha's followers, insists that Buddha tell him the answer to four questions. One of these questions is whether "the soul is the same as the body" or "the soul is one thing and the body another." Buddha's response to all four questions is to use the simile of a man struck by a poisoned arrow who refuses to pull it out unless certain questions are answered; after which Buddha concludes that he does not think it useful to answer any of these questions, since it will not help anyone overcome suffering and attain *Nirvāṇa*. So, in the spirit of the Buddha, and while I would acknowledge that the question of the relation of the mind to the brain can be philosophically interesting, I also have no need to try to answer it here. In addition, it is not clear to me that Metzinger's analysis of the PSM, whatever his metaphysical commitments, requires that the analysis itself presupposes the identity of mind and brain. The Phenomenal Self Model that Metzinger employs may be, as he argues, generated by brain processes, but his description of them is phenomenological. When, for example, he first introduces the PSM, he does so with a phenomenological description of the rubber hand illusion. He uses this illusion to demonstrate how the physical boundaries of what we take to be self can change. He then proceeds to discuss

¹¹ One way of interpreting Madhyamaka Buddhism's insistence on the emptiness of all phenomena is that there is nothing that is metaphysically basic, including even emptiness.

other ways in which the phenomenal self illusion is experienced. In fact, were he not able to provide such phenomenological descriptions, there could be no way of correlating the PSM with brain processes. For my purposes, the PSM is useful in helping us to understand some aspect of how the I-sense is constructed. That the PSM is causally connected to neural processes does not mean that it is only or “merely” the result of these processes or that it is nothing but these neural processes. Finally, suppose physicalist/identity metaphysics was correct. That alone would not establish that the self is an illusion, since everything we experience, including our perceptions, feelings, intuitions, and beliefs, would then be neural processes as well. There would still be a difference between illusion and veridical perceptions, between feelings that something is happening and whether it actually is happening, and between beliefs which are true and those which are false. In other words, the phenomenal sense of self would be a neural process whether or not there was a self.

Finally, I do not think that there is a necessary contradiction between the Buddhist analysis of the illusion of self and Metzinger’s analysis, because they are answering different questions. Buddhism does not reduce the *self* to the aggregates. What is reduced is the *person* to the aggregates. This reductionism does not explain how the I-sense is constructed, whereas Metzinger’s analysis of the PSM is an attempt to do this.

My use of evolutionary psychology raises some additional questions. Smith presents several objections to my assumption that the phenomenal illusion of self, as a result of natural selection processes, must have had and still has a useful function for the human organism. His first objection is that if the PSM has causal efficacy, it ought to be considered real from a Buddhist perspective, since, for Buddhism, “having causal efficacy marks it as real.” I think this objection confuses ultimate and conventional reality. For Buddhism, causal efficacy is ultimately the property of the *dharma*s resulting in the coming to be of other *dharma*s. But on the conventional level, we can still say that when persons believe certain things, this will cause them to act in certain ways; and that, on the basis of certain feelings and perceptions, they may come to have certain beliefs. Thus, if someone is thirsty and sees a mirage of a lake in the desert, this illusion will lead that person to run in that direction. On the level of ultimate reality, all this can be explained as the way that certain *dharma*s bring about other *dharma*s. To return to the problem of the PSM having causal efficacy, the I-sense on the level of ultimate reality consists of certain *dharma*s that will bring about other *dharma*s.

Smith’s second objection is to argue that my use of Metzinger’s PSM suggests that the function of the I-sense is to be a central causal agent. As against this, Smith refers to Siderits’ analysis of the way in which control shifts from one part of the psycho-physical processes to another, and that no one part is always the controller. This misses the point of saying that the illusion of self has a useful function. Its function is not to act as an agent who has executive control. Quite the contrary, what Siderits’ analysis demonstrates is that we can explain the *feeling* of agency without there being an agent who has executive control. What has a useful function is not the phenomenal self *as an agent* but the phenomenal *illusion* of being an agent. What is useful is not that there is a phenomenal self in charge, because there is no one who is in charge. What is useful is the *illusion* that there is someone in charge.

Finally, Smith objects to my use of Robert Wright's evolutionary psychology, because evolutionary psychology assumes that "our minds are the results of selections that have shaped our cognitive habits." This Smith insists is at odds with the Buddhist path which would promote "seeing things as they really are." There are several issues packed into this objection. First, evolutionary psychology makes in one sense a stronger claim – that certain features of both our cognitive *and* affective habits are the result of selection pressures. Second, as I will explain below, the claims made by evolutionary psychology are not as strong as Smith suggests, as at least some its proponents recognize that there is a significant interaction between our genetic inheritance and social factors in the formation of our cognitive and affective habits. Finally, Robert Wright does specifically believe that although selective pressures have shaped the illusion of self, Buddhism has resources to transcend this genetic inheritance; and I have suggested in my article that through *vipassanā* meditation *in conjunction with other components of the 8-fold path* it is possible to do this. Put simply, the human mind not only has forms of illusions built into it but also has capacities to "see things as they really are" and, therefore to escape from at least some of the "influence of our genetic programming."

At this point I need to confront two substantial criticisms that Siderits levels against my use of evolutionary psychology. The first involves a general criticism of evolutionary psychology, because it presupposes that "the brain is organized around a number of heritable mental modules." This, Siderits argues, is challenged by our understanding of the plasticity of the human brain. When one part of the brain is damaged, its tasks are often taken up by other parts. Thus, while allowing that certain functions of neural organization might be best explained as the straightforward result of natural selection – e.g., certain universal features of our visual system – it is highly unlikely that the phenomenal illusion of self could be generated by "a single heritable feature of brain organization." This is because the phenomenal illusion of self serves a number of functions, and that would present a "serious engineering challenge," given that it would involve "a large number of systems" responding to a "large variety of relevant stimulus conditions." I have several responses to this general criticism of evolutionary psychology and the possibility of it helping us to understand why the illusion of self developed. First, while evolutionary psychology has in much of its early formulations relied on the module metaphor, the metaphor is now the subject of intense debate within the discipline, and many of those who continue to use it also recognize the plasticity of the brain and distance themselves from the attempt to take metaphor too literally.¹² Secondly, Wright himself rejects many of the standard (mis)

¹² The idea of an evolutionary psychology goes back to Darwin who suggested that mental faculties may be the result of both natural and sexual selection. In contemporary evolutionary psychology, the initial assumption behind the modular metaphor was that psychological mechanisms that had evolved by natural selection to solve problems confronted in our evolutionary past were domain-specific. However, it is now widely accepted by many, if not most, evolutionary psychologists that the brain may be a combination of both domain-specific and domain-general features. Furthermore, that some psychological mechanisms are domain-specific refers to their ability to solve specific problems and does not necessarily entail that they exist in one compartmentalized site within the brain. In fact, many

interpretations of the modular metaphor and suggests that it might be better to think of them as networks or systems.¹³ Third, it is not necessary to think of the phenomenal illusion of self as hard-wired into the brain in order to see the relevance of evolutionary psychology. It is possible to understand what is heritable not as a specific set of modules but as a varied set of genes that develop the phenomenal illusion of self in conjunction with a variety of social conditions; and that this development permeates a number of the neural systems; in which case, it is not difficult to understand how one part of the nervous system might take over certain functions of the phenomenal sense of self when other parts are damaged. Finally, that the development of the I-sense at the phenomenal level might represent ‘a serious engineering challenge’ is not necessarily a telling blow against it. Evolutionary development is so full of solutions to serious “engineering challenges” that earlier philosophical theists like William Paley and contemporary philosophers like Anthony Flew, take them as evidence that there is some Super-Intelligent Designer who designed them.¹⁴ The answer by evolutionary theorists like Richard Dawkins to such theistic claims is that the grid of natural selection is sufficient to account for these solutions to complex design problems.¹⁵ Thus, even though the illusion of self performs a number of functions and even *if* it were solely based on a modular system in the brain, it would not from the evolutionary point of view pose a “serious engineering challenge.” I do not, however, think that the phenomenal illusion of self can be reduced to neural networks in the brain, and this leads me to a discussion of Siderits’ second criticism of my use of evolutionary psychology.

I have already granted in the previous section that the phenomenal illusion of self is only one of the components of the I-sense, the other being what Siderits calls the “bundle of cognitive and affective habits that continuously reinscribe belief in the self,” and I have added that these components interpenetrate and mutually reinforce each other. I now want to consider Siderits’ second criticism which, as I will explain below,

evolutionary psychologists allow that genes do not specify neural wiring in detail and recognize that neural networks are interconnected and subject to environmental influences. See Krill 2007 and Bolhuis 2011.

¹³ Specifically, he rejects the conceptualization of modules “as a bunch of physical compartments” located in specific parts of the brain, because brain scans reveal that a module may draw “on various regions of the brain, sometimes drawing more on one region, sometimes more on another;” he rejects the idea that modules are “like the blades on a Swiss army knife or the apps on a smart phone,” since there is “much more interaction among and overlap between different mental modules; and he rejects the interpretation of modules as like “departments in a company’s organization chart,” since modules may compete as well as collaborate, and there is no hierarchy among the modules (Wright 2017, 87-88).

¹⁴ Paley compared the human eye to a camera, and Antony Flew became a Deist after being a famous philosophical agnostic for most of his career. He was first led to question his agnosticism on the basis of reflections on the complexity of the double helix. See Flew 2007.

¹⁵ “What makes natural selection succeed as a solution to the problem of improbability, where chance and design both fail at the starting gate? The answer is that natural selection is a cumulative process, which breaks the problem of improbability up into small pieces. Each of the small pieces is slightly improbable, but not prohibitively so” (Dawkins 2008, 147). The result is that the accumulation of small pieces can create a more complex structure that only seems highly improbable, because it is thought of as occurring all at once.

pushes me to clarify and further amend my position, while still insisting that the phenomenal illusion of self has significant biological evolutionary roots.

While allowing that there are some psychological aspects present in the human brain at birth – e.g., a tendency to gaze more fixedly at smiling faces and gaze detection – Siderits argues that it is not until much later in the child’s life, approximately when the child is four years old, that the sense of self emerges. This is because it is not until the child reaches this age that it is able to fully develop “a theory of mind that is first applied to others and only then applied to oneself.” On this basis, Siderits argues that “insofar as the older child has been subjected to socialization in its home culture,” the I-sense would be predominantly the product of that socialization. From this, Siderits draws the conclusion that the I-sense is primarily the result not of biological evolution but of cultural evolution, “not of natural selection but of cultural selection.”

This strikes me as an odd conclusion. It doesn’t follow that because the sense of self emerges from interaction with others that it is the result of cultural selection. Cultures are incredibly different, and what we need to explain is the universality of the I-sense. Siderits’ only explanation for this universality is that cultures which did not have the sense of self did not survive; hence, cultural selection. But this assumes that there were a variety of human cultures which did not have an I-sense and, therefore, proved not to be successful. But if the I-sense is a result of applying a theory of mind to the other and then applying it to oneself, what would be the situation of child rearing in which that would not happen. Of course, there is the legendary idea of the “wolf child” who survives without early human interaction and does not develop a sense of self, but then wolf children are not going to initiate a human culture. Thus, it is unlikely that without an I-sense a human culture would ever come to exist in the first place. Still, I think there is one part of Siderits’ discussion of the cultural evolution of the I-sense that I think is correct. Different cultures do organize the I-sense in different ways. Clearly, the hyper-individualist I-sense of contemporary American culture is not the norm, even for all cultures existing today, and there is good reason to think that the I-sense throughout history was organized in quite different ways. The sense of self in European feudal culture was certainly markedly different than the sense of self in modern capitalist culture; and, given its Confucian underpinnings, the sense of self in Chinese feudal culture would be different from its form in European feudal culture.

Now I would agree that the sense of self is constructed through interaction with parents or other caretakers in the first several years of the child’s life, and I explicitly made use of George Herbert Mead’s analysis to make this point in my article. While Siderits recognizes that I did this, he points out in a footnote that I sometimes seem to be saying that the “phenomenal self is innate but only becomes manifest through normal developmental processes.” As I read over my article with this in mind, I must admit that I did indeed say several things that would lend itself to this interpretation, and I did not offer enough discussion of the way in which the innate components that give rise to the sense of self interact with the later social elements that construct it. So let me clarify by using Chomsky’s analysis of language as an analogy. Chomsky claims that we have an innate universal grammar which constrains the forms that language can take and makes possible both the development of many different kinds of languages

and the creative use of language.¹⁶ This universal grammar is then elicited by the child's exposure to the specific language (or sometimes more than one language) of its social environment, but the various world languages, even though they all obey the rules of the universal grammar, do exhibit differences that are significant. In short, the innate universal grammar is by no means the *whole of language* lying dormant in the infant at birth, simply waiting to be elicited. By analogy, I would suggest that what is innate is a set of neural connections which make it possible for the child to interact with others and, in the first few years of life, to apply a theory of mind to others and then to itself or, to use Mead's analysis, to see itself through the vantage point of other specific persons and eventually from the vantage point of the group as a whole. These interactions do elicit the core of the sense of self but not the whole of it. Eventually, as children learn the language of their culture, they also learn various linguistic activities ('language games'), including the use of indexicals; they internalize those activities and – given the tendency to essentialize and wish to endure – further develop their sense of self; and then through narrative construction this culminates in the full development of the sense of self – the autobiographical self. In other words, the sense of self does not exist in a fully developed latent form at birth, simply ready to be turned on by the social environment any more than French, English, Chinese, or Japanese exist already in a latent form at birth. What exists innately are certain neural components which generate certain kinds of interaction and an ability to draw certain conclusions from those interactions (e.g., ability to know the mind of the other, sometimes called "mind reading"); *and* the inevitability that, *under normal social conditions*, the child will do so. From there, the sense of self develops its specific cultural form, as the child is socialized by the language, literature, media, and norms of its culture.

6. VIPASSANĀ, EXTINGUISHING THE I-SENSE, AND THE BUDDHIST PATH

I now need to consider two other criticisms raised by Smith and Siderits, respectively. In my article, I emphasized *vipassanā* meditation as a method for seeing through the illusion of self, for verifying the main assumptions of the arguments for no-self, and, *in conjunction with other components of the 8-fold path*, as a way of extinguishing the phenomenal illusion of self and developing a sense of selflessness, a non-self perspective. However, I further argued that, even as one approaches enlightenment and

¹⁶ In an interview with Kevin Doyle, Chomsky talked about the relation of human nature to language: "There is nothing *regressive* about the fact that a human embryo is so constrained that it does not grow wings, or that its visual system cannot function in the manner of an insect, or that it lacks the homing instinct of pigeons. The same factors that constrain the organism's development also enable it to attain a rich, complex, and highly articulated structure....Take language, one of the few distinctive human capacities about which much is known. We have very strong reasons to believe that all possible human languages are very similar; a Martian scientist observing humans might conclude that there is just a single language, with minor variants. The reason is that the *particular aspect of human nature* [italics mine] that underlies the growth of language allows very restricted options.....It is these very restrictions that make it possible for a rich and intricate system of expression of thought to develop in similar ways on the basis of very rudimentary, scattered, and varied experience" (Chomsky 2001, 20).

perhaps even when enlightened, it may not be possible to maintain the non-self perspective consistently. Smith poses a challenge to my claim about the epistemic role of *vipassanā* meditation as a form of “verificationism” and suggests that it might have a different function in Buddhist soteriology. Siderits argues, as I briefly mentioned earlier, that extinguishing the I-sense requires extinguishing the bundle of cognitive and affective habits that constitute the I-sense and this requires not extinguishing the phenomenal illusion of self through *vipassanā* but the creation of new cognitive and affective habits.

I will begin with Smith’s challenge to my understanding of the epistemic role of *vipassanā*. Smith argues that “we should not think of mindfulness meditation as a way of verifying Buddhist philosophical insights but as a way of actively conditioning the perceptual apparatus such that those Buddhist principles and views we antecedently endorse become pre-reflective ways of taking up our world.” He offers as an example a form of mindfulness meditation wherein the practitioner is encouraged to imagine the body as a corpse in the various stages of decay or as a set of dismembered elements, the way a butcher might look a cow. The point of these mental exercises, Smith argues, is not verification but to project an image onto the body whose purpose is to internalize an already held philosophical position and create an attitude of detachment. Finally, Smith also argues that thinking of *vipassanā* as a form of verificationism “flattens the normative implications of the practice.”

I will begin my response to these worries by focusing on the examples he gives of an alternative form of mindfulness. Clearly this is a different form than the non-judgmental moment to moment attention that I discuss in my article. Still, the difference is less significant than it might initially appear. Both may be considered forms of controlled experiments, and the thought experiment which Smith highlights has as its purpose to focus on something that one can recognize is true about the body – that since it will decay or can be carved up like a piece of meat, it is a mistake to become attached to it. While these thought experiments begin with projection and imagination, it is fundamentally connected to something real, unlike a thought experiment where one imagines oneself, for example, as a butterfly. The thought experiment can then function as a hypothesis which can be confirmed by focusing on what happens to one’s attitude to the body if one performs the experiment in a serious and consistent way. I think it can be verified that individuals who embark on this experimental practice, will tend to develop an attitude of non-attachment to the body.¹⁷

The form of *vipassanā* that I discuss in my article – an open receptivity and non-judgmental attention to passing phenomena – is also a form of controlled experiment. It is not the way we “normally” function in the world, and it is not like ordinary introspection. As I noted in my article, one Buddhist practitioner and teacher of mindfulness compares it to an electronic microscope. I think this metaphor is especially appropriate, as this method of meditation allows one to observe the processes of the

¹⁷ The experiment can be performed more vividly and realistically. Monks in Thailand still to this day go to a morgue and observe actual corpses in various stages of decay, saying to themselves “this is my fate.”

mind in micro-detail. In this way it can verify certain assumptions that Buddhism uses to argue for the doctrine of no-self – the impermanence of these processes, that they arise in dependence on other processes, that they are not subject to control, etc. Nonetheless, I would, agree with Smith that the main point of this attention to the continuous flow of psychophysical process is to internalize philosophical assumptions “antecedently endorsed” and to recondition our faculties in order to bring about a different way of being in the world. Thus, the experiential verification of the Buddhist assumptions is a verification of how those principles can be applied to and re-condition one’s own experience. Eventually, it begins to verify that the self is an illusion through micro no-self experiences (micro-moments of *Nirvāṇa*) which, as one continues on the path, become more predominant. Thus, the process of verifying Buddhist principles does not flatten the normative and soteriological implications of the Buddhist practice but rather serves as one of the initiators of the Buddhist path as well as one of its continuous components.

Still, there is the problem of what it means to “verify” these assumptions, since these assumptions guide the process of *vipassanā* meditation. In other words, the verification process seems to be relying on the very principles that are to be verified. There is, thus, no possibility of falsifying those principles. I take this as a serious question whose answer would require much more than I can say here. What I will say briefly is that strict Popperian falsifiability is not the only way to understand verification. To understand how verification might work in *vipassanā* practice and, indeed, in the whole of the Buddhist path, we might think of Buddhism as a “research program” in Imre Lakatos’ sense of that term.¹⁸ Understood in this way, the Buddhist research program would not be judged by whether its core principles were falsifiable, but whether embarking on and following the Buddhist path can lead to progress toward its soteriological goal, whether “walking the walk” based on experientially knowing and internalizing Buddhist principles leads to the progressive overcoming of suffering.

In some ways I have already provided my answer to Siderits’ criticism, but I now want to look at his criticism in some detail and explain my response. I already discussed Siderits’ claim that the I-sense is not a phenomenal illusion but is constituted by a bundle of cognitive and affective habits. If this were all there was to the I-sense, it would then follow that to extinguish the I-sense is not to extinguish the phenomenal illusion of self but to extinguish these habits. The key to doing this, for Siderits, is not to dispel something with a phenomenal character but to change the habits of thought that produce delusion and to cultivate “counter virtues, affective habits that help us unlearn such deeply engrained vices as anger, resentment, and jealousy.” Thus,

¹⁸ See Lakatos 1970. For Lakatos, science is best understood as a research program consisting of a set of hard core theories, which are refutation resistant, and a variety of “auxiliary hypotheses,” which in conjunction with the hard core theories generate predictions that are subject to empirical tests. If the predictions are disconfirmed, it is the auxiliary hypotheses that are given up, while the hard core is retained. The research project can continue so long as it is “progressive,” which is to say, as long as it can predict new and unexpected facts that can be tested and as long as it is capable of solving certain anomalies that previous research programs cannot solve.

extinguishing the I-sense would require not *vipassanā* but meditative practices that can cultivate these counter-virtues.

Siderits highlights sympathetic joy and loving kindness as counter-virtues to be cultivated. These are two of the four “immeasurables” or “boundless” qualities known as the *brahma viharas*, the other two being compassion and equanimity. Early Buddhism and the Theravāda tradition have developed elaborate meditative techniques to cultivate these qualities as has various Mahayana traditions.¹⁹ Thus, none of the Buddhist traditions would disagree with at least one part of Siderits claim – that overcoming our self-orientation requires forms of meditation that develop new affective habits through the cultivation of counter-virtues. However, if there is, as I have argued, a phenomenal illusion of self that is at least a component of the I-sense, then it would seem to be appropriate to approach the problem of the illusion of self in several ways – to extinguish it as a phenomenal illusion through *vipassanā* and to extinguish it a set of affective habits through meditative techniques that can cultivate the *brahma viharas*. As for the cognitive habits, it would seem appropriate to confront them with philosophical reflections *and* with both forms of meditation, as they are bound up with the phenomenal illusion of self and the affective components of the I-sense. In all, since the phenomenal illusion of self and the cognitive and affective components of the I-sense intersect and interpenetrate, it is important to disentangle the knot at all its corners.

To conclude, let me return to the analogy I used earlier. If we take the I-sense as a program which organizes our experiences into forms of self-cherishing and self-grasping, a program which produces the phenomenal illusion of self and develops self-oriented cognitive and affective habits, then we need to dismantle the operating program through both forms of meditation as well as through the other components of the eight-fold path. And as we dismantle our self-cherishing and self-grasping program, we can begin to put in its place a new program of selfless loving kindness, compassion, sympathetic joy, and equanimity.

Once again, let me thank Anand Vaidya, Sean Smith, and Mark Siderits for their insightful comments and provocative criticisms. They have helped me think through some of the issues at greater length, pushed me to further clarify and develop my position and arguments, and, in some cases, to revise them. I hope my response here will serve to further continue the dialogue in the future. Namaste.

¹⁹ For an excellent discussion of the significance of the *brahma viharas* and meditative techniques to cultivate them, see Feldman 2017. The Mahāyāna traditions tend to emphasize forms of meditation that cultivate boundless compassion. The Tibetan Buddhist tradition has developed a set of meditative practices known as *lojong* to get “the practitioner to adopt the project of *bodhicitta* in place of the project of self-cherishing” (Locke 2018, 262).

REFERENCES

Abbreviation

- MN: *Majjhima Nikāya*, trans. Bhikkhu Bodhi (2009), *The Middle Length Discourses of the Buddha: A Translation of the Majjhima Nikāya* (Somerville: Wisdom Publications).
- Bolhuis, Johan J., Brown, Gillian R., Richardson, Robert C., and Laland, Kevin N. (2011), “Darwin in Mind: New Opportunities for Evolutionary Psychology”, *PLOS Biology* 9.7: 1-8. <<https://doi.org/10.1371/journal.pbio.1001109>>
- Chomsky, Noam (2001), “Noam Chomsky on Anarchism, Marxism, and the Hope for the Future,” *Red and Black Revolution* 2: 17-21.
- Dawkins, Richard (2008), *The God Delusion* (New York: Houghton Mifflin Company).
- Engler, Jack (2003), “Being Somebody and Being Nobody”, in Jeremy D. Safran (ed.), *Psychoanalysis and Buddhism: An Unfolding Dialogue* (Boston: Wisdom Publications), 35-100.
- Feldman, Christina (2017), *Boundless Heart: The Buddha’s Path of Kindness, Compassion, Joy, and Equanimity* (Boulder, Colorado: Shambhala Publications).
- Flew, Antony with Varghese, Roy Abraham (2007), *There is a God: How the World’s Most Notorious Atheist Changed His Mind* (New York: Harper Collins).
- Garfield, Jay; Nichols, Shaun; and Strohming, Nichols (2018), “Episodic Memory and Oneness”, in Philip J. Ivanhoe, Owen J. Flanagan, Victoria S. Harrison, Hagop Sarkissian, and Eric Schwitzgebel (eds.), *The Oneness Hypothesis: Beyond the Boundary of Self* (New York: Columbia University Press), 285-304.
- Gowans, Christopher W. (2003), *Philosophy of the Buddha* (New York: Routledge).
- Krill, Austen L., Platek, Steven M., Goetz, Aaron T., and Shackelford, Todd K. (2007), “Where Evolutionary Psychology Meets Cognitive Neuroscience: A précis to Evolutionary Cognitive Neuroscience”, *Evolutionary Psychology* 5.1: 232-256. <<https://doi.org/10.1177/147470490700500114>>
- Locke, Jessica (2018), “Training the Mind and Transforming Your World: Moral Phenomenology in the Tibetan Buddhist *Lojong* Tradition”, *Comparative and Continental Philosophy* 10.4: 251-263. <<https://doi.org/10.1080/17570638.2018.1531468>>
- Lakatos, Imre (1970), “Falsification and the Methodology of Scientific Research Programmes”, in Imre Lakatos and Alan Musgrave (eds), *Criticism and the Growth of Knowledge* (New York: Cambridge University Press), 91-196.
- Marx, Karl (1994), “Toward a Critique of Hegel’s *Philosophy of Right*: Introduction”, in Lawrence H. Simon (ed.), *Karl Marx: Selected Writings* (Indianapolis, Indiana: Hackett Publishing Company), 27-39.
- Mead, George H. (1963), in Charles W. Morris (ed.), *Mind, Self, and Society* (Chicago: The University of Chicago Press).

- Metzinger, Thomas (2009), *The Ego Tunnel: The Science of the Mind and the Myth of the Self* (New York: Basic Books).
- Siderits, Mark (2007), *Buddhism as Philosophy: An Introduction* (Indianapolis, Indiana: Hackett Publishing Company).
- Struhl, Karsten J. (2018), “Buddhism and the Problem of Universal Compassion,” in Justin Caouette and Carolyn Price (eds.), *The Moral Psychology of Compassion* (New York: Rowman and Littlefield).
- Wikipedia, “The Illusion of Control”. <https://en.wikipedia.org/wiki/Illusion_of_control>. accessed November 15, 2020.
- Wright, Robert (2017), *Why Buddhism is True: The Science and Philosophy of Meditation and Enlightenment* (New York: Simon and Schuster).