CONSTRUCTIVE-ENGAGEMENT DIALOGUE

PHILOSOPHY OF MIND

AUTHOR MEETS CRITICS:

BORN BELIEVERS?

MARK SIDERITS

In “What Kind of an Illusion is the Illusion of Self”, Karsten Struhl brings up an important question, and develops an interesting answer. There is much here that I agree with. There are, though, two questions raised by his paper that I believe may warrant further investigation. These are: (1) is the self a phenomenal illusion? and if so, (2) is its being a feature of human psychology a straightforward result of natural selection? How the second question should be addressed might seem to depend on how the first is answered. But in part because it will prove difficult to work out just how to approach (1), I shall start with (2).

To say that the self is a phenomenal illusion is to say at least this much: that something that is readily mistaken as evidence for the existence of a self commonly appears in the experience of ordinary humans. Let us suppose we understand what this comes to—perhaps that there is a raw “what-it-is-like”-ness to our experience that gives rise to the belief that we have selves. At least one such belief is that the experiences presently occurring are had by a subject. So if the self is a phenomenal illusion, then ordinary sensory experience would be given to us in a way that leads us to spontaneously affirm that they are given to or for a particular experiencing subject, me. Let us similarly suppose that features of our experience render it difficult not to judge that this me is an enduring substance located in my head, something that serves to ground certain synchronic and diachronic unities and is the author of my actions. Let us suppose, that is, that the standard characterization of a self—as the persisting substance that grounds the synchronous and diachronic identity of the person in the guise of experiencing subject and agent—is supported, albeit misleadingly, by the felt

SIDERITS, MARK: Emeritus Professor of Philosophy, Illinois State University, USA. Email: msideri@ilstu.edu

1 Struhl suggests that the notion of a phenomenal self should be understood in this way when he uses such expressions as ‘the phenomenal appearance of the self’ (136).
character of our experience. To suppose that the experiences of humans ordinarily have this phenomenal character might be akin to what Buddhists claim when they say that the “I”-sense (āhamkāra) is a human universal, something common to all unenlightened humans (and perhaps to non-human sentient beings as well).

Buddhists explain the universality of the “I”-sense by calling it the result of “beginningless ignorance.” Their explanation builds on the assumption that the series of rebirths of which one is the latest instance has no beginning. Since each life in such a series resulted from intentional actions performed under the sway of ignorance in its predecessor life, it would then follow that the illusory “I”-sense has always been with us. For those who do not believe in a beginningless series of rebirths but do think that the self illusion is rooted in our phenomenology, there may be other ways to explain how things appear to us. Those of a naturalistic bent will be particularly tempted to bring selectionist machinery to bear on the question. But which sort of selection, natural or cultural? If the “I”-sense derives from a phenomenal character of experience that is widely shared among humans, is this because a nervous system so organized as to yield this character was selected for in the environment of evolutionary adaptation? Or is it rather that the human genome merely facilitated the development of human culture, and that among the various resulting cultures, those that promoted the development of an “I”-sense had greater long-term success?

There are passages in Struhl’s paper that appear to credit both types of process with a role, e.g., “From an evolutionary point of view (both biological and cultural evolution) the illusion of self was a necessary illusion” (128). But then there are also passages like this: “natural selection built into our species the illusion of self” (129). And this: “their brains are still organized by the imperatives of that genome…. The way the human brain is naturally organized to do this employs the device of the phenomenal self” (135). Such passages suggest that culture plays at best an auxiliary role, perhaps one of merely facilitating the manifestation of traits already programmed into the nervous system by the human genome. Now if the self is a phenomenal illusion, then knowing its cause may matter to how we are to address it. Are we born believers in a self, or does that belief depend on a particular process of enculturation?

It may be no accident that the passages where Struhl seems to favor a natural-selection explanation are also ones that invoke the support of the evolutionary psychologist Richard Wright (2017). The “just-so” stories of evolutionary psychology rely on the idea that processes of natural selection led to a human brain that is configured to predispose us to think and act in certain predictable ways. In the present

---

2 To say this is not the same thing as saying that the self is ‘socially constructed’. What is to be explained here is that the illusion of a self is a human universal. A cultural selectionist account of some phenomenon acknowledges the diversity of human cultures, but explains the widespread sharing of the phenomenon across different cultures by claiming that those early human cultures in which the phenomenon occurred had a competitive advantage over cultures lacking that feature. That the phenomenon is now to be found across many distinct cultures is not explained by facts about human biology, but rather by facts concerning what drives competition among human cultures. To say that this is true of the self illusion is perfectly compatible with the thought that the self may be socially constructed in quite different ways in different cultures.
context the idea would then be that our brains are hard-wired to generate experiences of such a character as to produce a pre-reflective “I”-sense, and this is explained by the enhanced reproductive fitness that sense conferred on our ancestors in the environment of evolutionary adaptation. But as always with the explanations proffered by evolutionary psychology, this account requires that we suppose the brain is organized around a number of heritable mental modules. And this seems implausible given what we now know about the brain’s plasticity: damage to one part of the brain is often followed by other parts of the brain taking up tasks formerly performed in the damaged region. A plausible natural-selectionist explanation can be developed for the Kanizsa square illusion. The human visual system computes shapes through detection first of color and then of edges, with the latter computation relying on color-shading gradients. The system deals with cases where borders are only partially delineated by inducing brightness and thereby producing regions of heightened contrast. And so the square one seems to see appears brighter than the encompassing white. That the illusion is universal can be explained by the relative success conferred, by a vision system so organized, on those of our animal ancestors in whom such a system first occurred (through a series of transcription errors). Shape detection is clearly important to the success of sighted animals, and this configuration works well enough in environments typical of the environment of evolutionary adaptation. Much of the neural organization involved is actually not in the visual cortex of the brain but on the retina. So the hypothesis that such organization might be heritable is quite plausible. Far less so the hypothesis that there is a single heritable feature of brain organization responsible for generating the phenomenal illusion of an “I.” For think of all the tasks this illusory appearance is meant to perform: grounding both synchronic and diachronic unifications of several different sorts, serving as subject of experience, serving as agent of action, etc. The large number of different systems that would be necessary to generate a phenomenal “I”-sense in the large variety of relevant stimulus conditions represents a serious engineering challenge. Add in the plasticity of the human brain and the odds against an illusory phenomenal self being the result of natural selection seem quite daunting.3

To say this is not to say that there is nothing in the human brain at birth that might help explain the development of a phenomenal “I”-sense. Here are some examples of infant behaviors that might well be innate. Neonates look intently at highly stylized pictures of smiling faces. Infants are quite good at detecting the direction at which another person is looking. And they use this ability to try to coordinate attention to objects in the shared environment—to bring about joint attention with possible caregivers. It seems reasonable to suppose that each of these is subserved by some feature of the infant brain. Together they and certain other capacities constitute what Bogdan (2010, 31) calls the infant’s “naïve psychology,” something that will over time

---

3 Indeed Struhl quotes (125) Hood to this effect: ‘The brain has many distributed jobs…. [T]he sense of self that most of us experience is not to be found in any one area. Rather it emerges out of the orchestra of different brain processes like a symphony of the self, just as Buddha and Hume said’ (Hood 2012, xii).
be replaced by the full-fledged theory that he calls “common-sense psychology.” And it is not difficult to fashion natural-selection explanations for these and other individual components of the infant’s naïve psychology. The neonate’s response to the “smiley-face” stimulus, for instance, will clearly play a role in fostering infant-caregiver bonding, particularly when, starting a few days after birth, the infant begins to smile in response to the smiles of others. Gaze detection is likewise crucial to achieving coordinated or joint attention, something that will help in satisfaction of the infant’s nurturance needs given its reliance on adult caregivers. These capacities might be thought of as forming the rudiments of a theory of mind. But caution is called for here, since it would be unjustifiable to attribute to the infant anything as sophisticated as the construction and testing of hypotheses concerning the springs of human behavior. All of that will, on Bogdan’s account, come much later in the developmental story, with the achievement of what he calls common-sense psychology in later childhood. The important point here is that each component behavior can be understood as involving a neural mechanism that was selected for in the environment of evolutionary adaptation.

The situation of the human infant is, after all, quite dire, given that it is the outcome of the obstetrical dilemma posed by human bipedalism. The narrowing of the birth canal that was required for live birth in a bipedal mammal has made humans an extremely altricial species; infant survival depends entirely on adult caregiving. (It has also meant that live birth requires the cooperation of other adults besides the mother, so a high degree of sociality is also necessary for survival of the species.) Consequently, if a tendency to gaze more fixedly at smiling faces were a heritable feature of the human brain, then the spontaneous appearance of this feature in the brains of certain of our ancestors would have given them some advantage in having their nurturance needs met, thereby accounting for its prevalence in the species today. And the behavior involves a sufficiently stereotyped stimulus that it is not difficult to imagine its being tied to a particular brain structure. A heritable mental module for paying differential attention to smiling faces does not seem all that far-fetched.

It is also plausible that without these components of “naïve psychology,” the child would not develop (typically beginning at around age four) the full-fledged theory of mind that constitutes our adult common-sense psychology. But, argue Bogdan and others, it is only with the development of this theory that the child comes to attribute mental states to itself. And this seems to mark the point at which the child can meaningfully be said to have a sense of self. The achievement of a sense of self follows on the development of a theory of mind that is first applied to others and only then applied to oneself. This would suggest that the phenomenal self could only appear later in childhood and not at birth. And insofar as the older child has been subjected to socialization in its home culture, it becomes more plausible that a phenomenal self is the product of socialization in a particular culture. Now Struhl has discussed the various

---

4 Struhl seems to agree: “There is good reason to think that the phenomenal sense of self was initially constructed in the early stages of the child’s life through interaction with parents or other caretakers” (123). But then there is, “it is elicited and turned on through our early childhood interactions with caretakers” (132), suggesting that his phenomenal self is innate but only becomes manifest through normal developmental processes.
benefits that come to individuals and their society from their having an “I”-sense. But cultures vary across space and across time, and such variation sets the stage for cultural evolution. So it is plausible that if there is a phenomenal self, it is the result not of natural selection but of cultural selection. Ordinary people have the “I”-sense not because they were born with it, but because they were socialized in one of the cultures that developed strategies for inculcating such a sense—and that thereby proved more successful than those early human cultures that did not. Like human language, an “I”-sense would have arisen in the presence of certain biological precursors, but only because of processes of cultural selection.

This result, however, makes it appear less likely that the self can be a phenomenal illusion. Recall that we began with two questions: (1) is the self a phenomenal illusion? and if so, (2) is its being a feature of human psychology a straightforward result of natural selection? Our negative answer to the second question makes a negative answer to the first more likely. To see why, we must say more about what it might mean to call the phenomenal self an illusion. A phenomenal illusion, we are told, is a phenomenal presentation brought about by the way our mind naturally structures our experience and that strongly inclines us to hold a false belief. Struhl uses the Kanizsa square illusion as his example, but there are many others, including most famously the Müller-Lyer illusion. Buddhist discussions of illusions often cite examples involving malfunctioning sensory systems, as in the net of hairs seen by those with the ophthalmic disorder known as timira, the two moons seen by someone who is drunk or dizzy, or the conch shell seen as yellow by one with jaundice. While one might say that the mind structures the experiences involved, this is not how the mind naturally structures them. But Buddhists also cite mirages, the Fata Morgana or City of the Gandharvas, the circle of fire one sees when a torch is rapidly twirled, and the apparent motion of trees on the river bank seen by someone on a boat. One might also include here the vile smell that clean water presents to the properly functioning olfactory system of pretas. Unlike the net of hairs and the two moons, these all involve the normal functioning of sensory systems.

Struhl uses a cognitive penetrability test to distinguish between phenomenal illusions and what he calls cognitive illusions: if knowledge that the induced belief is false dispels the illusion, then it is cognitive, if not then it is phenomenal. The Müller-Lyer illusion is phenomenal because even after one has carefully measured the two lines and ascertained that they are the same length, one line continues to appear longer than the other. But does this test work for all cases? The case of the moving trees

---

5 Buddhists agree about the benefits. See Milinda Pañha 40.
6 For the parallel case that language use likewise has biological precursors but developed through processes of cultural evolution, see discussion in Thomas Kirby 2018.
7 In line with their disjunctivism, Buddhist epistemologists call these cases of pseudo-perception (pratyakṣabhāsa). Dignāga famously claims that these do not involve conceptual superimposition. This would go some way toward explaining their supposed cognitive impenetrability.
8 Rebirth as a preta comes from bad karma, the fruit of that karma being the hunger and thirst that result when their sensory systems make all food and drink revolting. The preta’s sensory systems are functioning properly when they make nourishing food smell and taste like feces.
illusion might give us pause. Here is an updated version: the inbound train one is on is stopped at a station, and an outbound train is stopped on the opposite track. One sees the other train depart, but a moment later kinesthetic feedback alerts one to the fact that the train one is on is moving. One no longer sees the other train moving. The illusory motion of the other train is dispelled by new information. A Buddhist epistemologist might object that motion is itself a conceptual superimposition, something not perceived but inferred: since everything is momentary, nothing exists long enough to move from one location to another. And if attribution of motion is conceptual, then it is no wonder that the illusory motion of the other train should prove to be cognitively penetrable. But then if our awareness of motion is not tied to a specific phenomenal character—if motion is something not perceived but inferred—we must begin to wonder just what phenomenal character amounts to.\footnote{Full disclosure: Galen Strawson listed me as among those who subscribe to what he calls the Great Silliness. (See “Magic, Illusions, and Zombies: An Exchange”, Daniel C. Dennett, reply by Galen Strawson, New York Review of Books, 3 April 2018.) This makes me something of a qualia-skeptic. My argument here, though, will be that even qualia-philes should reject the claim that the self illusion is like illusions that they think involve qualia.}

Such doubts are only amplified when we come to the case of the phenomenal self. 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. It is not just Hume, though, who fails to find a self when “looking within” and thus presumably investigating the phenomenal character of our experience. Take the case of the subject of experience that is supposedly presented when we reflect on our own experiential lives. The closest it seems we can get to giving phenomenal content to our conviction that there is such a subject is the sense that in addition to the intentional object of a particular experience, there is also the awareness that takes it as object. But this awareness turns out to be elusive: it is said to be something the awareness of which is necessarily non-thetic. This is the feature that led Brentano to characterize the consciousness component of an experience as \textit{nebenbei} or “by the way”—always glimpsed as something just receding over the horizon, something caught only out of the corner of the eye (Montague 2017, 363). And if the cognizing activity of the experiencing subject should be beyond our grasp, what does this say about the supposed author of this activity?

The belief that this subject of individual experiences is also something that endures is said to be induced at least in part by the phenomena of episodic and autobiographical memory. But what might be the phenomenal character of such rememberings that would support this belief? Consider a case where a present stimulus triggers a memory of an earlier experience: I see a mango in the market and remember feeling pleasure when I tasted my first mango years ago. I say, “I want that mango because it reminds me of the first time I ate one.” My report is in the first person, and it reflects an assumed diachronic personal identity. But what is there in the phenomenal character of my experience that could be said to make that assumption seem inescapable? To see the difficulty here, consider an impersonal formulation: present visual experience triggers a memory trace that is the product of a past gustatory experience in the causal series of
mental events associated with this brain and body, and this in turn causes the formation of a desire. The impersonal formulation obviously omits the “I” of the first-person report. And if we think of “I” as an indexical,\(^\text{10}\) we may wonder how the impersonal formulation can do the same work as the first-person formulation. Indexicals are necessary in order for information to guide action: the map is useless for navigation if one doesn’t know where here is. So how can the impersonal formulation rationalize my action of buying the mango? But the impersonal formulation does contain an indexical: “this.” The expression “the causal series of mental events associated with this brain and body” would seem to do all the same work, albeit in a more verbose way. What this should suggest is that there is no distinctive phenomenal character expressed by reports from the first-person perspective. The “I” is the classic case of a many masquerading as a one: the many strongly casually connected psychophysical elements masquerading as a single enduring substance.\(^\text{11}\) It is our habitual use of the convenient designator “I” that leads to the sense of diachronic identity. There is nothing in our mistake that is the analog of the white square we see when we look at the Kanizsa figure.

If the illusion of self is of the cognitive and not the phenomenal sort, then it should be cognitively penetrable. And Struhl is of course right that merely rehearsing the philosophical arguments against the existence of the self does not by itself extirpate the “I”-sense. By all accounts, Hume was not an arhat. But here it might be useful to consider what Abhidharma says about the matter. The goal of overcoming suffering by extinguishing the “I”-sense is achieved, they claim, by ridding oneself of the defilements (kleśas). And here they draw a distinction between those that are extinguished through wisdom (prajñā) alone, and those whose extinction requires deployment of techniques of meditation (bhāvanā). Among the former are the variety of views concerning an existent person (satkāyadṛṣṭi). So explicit belief in a self should be corrigible by way of a strictly theoretical exercise. Such views are classed under the category of delusion (moha), the third of the three defilements. But many of the dispositions classed under the other two defilements are more straightforwardly affective in nature, and the uprooting of these is said to require meditative practice. Typical of the defilements classed here are such reactive attitudes as anger, resentment and jealousy. To take a simple example, suppose that the mango I hanker after was just purchased by another shopper, whom I now see eating it with great delight. I feel jealousy. This response is predicated on the premise that my own pleasure is of more value than that of others. This in turn involves the presupposition that there is such a thing as what makes events in this causal series belong to me, to be mine. Even if we know this presupposition to be false, we might expect the attitude to nonetheless arise. And occurrences of this attitude reinforce the disposition, which in turn reinforces our tacit engagement with the “I”-sense. The meditative counter-measure recommended here is not mindfulness or concentration meditation, but the cultivation of counter-

\(^{10}\) For discussion of whether “me” is an indexical or instead a Millian name (a “mere tag”), see Echeverri forthcoming.

\(^{11}\) See Parfit 1984, 206 for “strong connectedness”.

Comparative Philosophy 12.1 (2021)
virtues, affective habits that help us unlearn such deeply engrained vices as anger, resentment and jealousy. Given that we have been socialized to construct narrative selves—for all the reasons that Struhl describes—and narrative selves are built around prudential concern and the happiness-seeking project, this will prove difficult. It’s hard to learn to take pleasure in the other shopper’s delight in that mango and not regret my missing out. But this is just what the meditative practices of sympathetic joy and loving-kindness are meant to instill.

The lesson to be learned from this is that extirpating the “I”-sense is not a matter of eliminating a certain phenomenal character from our experience of the world. The “I”-sense is not a delusive phenomenal character. “I’-sense” is 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. How the world appears to us is not changed when this sense is extirpated. What changes is just how we reflexively respond to the world. There may be a phenomenal square in the Kanizsa illusion, but there is little reason to think that there is a comparable phenomenal self involved in the illusion of self.

There are two further points I think might be made. The first concerns a third way in which Struhl thinks the self might be thought to be illusory, besides as phenomenal illusion and as cognitive illusion; this involves the Freudian notion of wish-fulfilment as a mechanism for generating illusions (121). He points to the fact that belief in an enduring self might help satisfy the wish that our lives have meaning, a wish that is in danger of being undermined by the fact of our transience. It is not clear that this mechanism can explain how the illusion of self is generated, however. One cannot sense that one’s happiness-seeking project is threatened by one’s mortality unless one already sees the events in this life as one’s own, as events that are experienced by and can have meaning for this self. The Freudian wish-fulfilment mechanism might explain resistance to evidence that there is no self, but it cannot, on pain of circularity, explain our coming to have the belief that there is such a self.

My second and final point about Struhl’s paper is this. Some will criticize Struhl’s use of results from cognitive science and evolutionary psychology to explicate Buddhist ideas. And it is true that the naturalism of such of the authors Struhl cites as Dennett, Hood, Metzinger and Wright is incompatible with the uniformly anti-physicalist stance of Indian Buddhist philosophy. But it is not clear to me why this should be thought to matter. The question I take Struhl to be addressing is what sort of illusion an illusory self might be. Buddhists agree that illusions can be generated by the operation of our sense faculties, as well as by ways in which our cognitive machinery characteristically operates. Buddhists did not try to explain those modes of operation by appealing to processes of evolution by natural or cultural selection. The question at issue here, though, is not the etiology of the illusion but the nature of the illusion, whether it can be dissipated, and if so how. Understanding the mechanisms that explain how we came to be subject to the illusion may be of interest, but such understanding is not necessarily crucial to answering the practical question. That classical Indian Buddhists have more in their ontology than do current physicalists seems neither here
nor there, a matter of ontological bookkeeping with no real bearing on the question of how the illusion is generated and how its effects might be overcome.²

Here is one last point I would like to make, in this case not about Struhl’s paper but about the paper of a fellow respondent. Sean Smith mentions, in his paper, a view that Struhl and I happen to share concerning the Buddhist goal of the cessation of suffering. The view is, roughly, that it is existential suffering that is the principal target of Buddhist practice, and that such suffering is connected to non-self by way of the point that when happiness is understood as involving a sense of purpose and fulfillment, our search for happiness will inevitably be frustrated, given our mortality. The realization of non-self is thus meant to dissolve the underlying presupposition that there is a “me” for whom events in this life might have meaning. The connection can be seen in the fact that young people are routinely told that they must figure out who they truly are if they are to work out what sort of life will make them happy—what sorts of projects will confer meaning and dignity on their lives. Smith disagrees with this understanding of the Buddhist path, claiming “that there is a notion of meaning and purposefulness in one’s life available to the Buddhist that does not entail the existence of a self.”³ It is not my place to respond on Struhl’s behalf, but I thought I might say a word about my own stance here.

Smith cites the famous “glass tunnel” passage (Parfit 1984, 281) where Parfit describes his own response after coming to believe that one’s continued existence does not involve a “further fact.” Smith understands Parfit to be saying that such a dissipation of the sense of self leads one to be more inclined to “serve others,” something Smith says he finds “affirming and worthy of pursuit.” If “serving others” means preventing the suffering of others, then Buddhists would surely agree that this is worthy of pursuit. As Prajñākaramati says, “If suffering is to be prevented, then all is to be prevented” (yadi vāryam duḥkham, tādā sarvam vāryam BCAP ad BCA 8.103). But affirming of what, exactly? When we call a pursuit affirming, this is usually taken to mean that the activity confers value on the person who engages in it. When I find myself wishing that I could act on Prajñākaramati’s advice and follow the bodhisattva path, it’s because I feel that doing so would make me a better person, namely by imparting a larger meaning to my life. We are thus brought back to the notion of self as both narrator and central character in the ongoing story of a life. But I have written more extensively about this elsewhere,⁴ and bring it up here only to underline the point that there may be reasons to exercise caution in ascribing to Buddhists the view that the goal of their practice is a source of meaning and happiness. There may be a reason

—

² Unless, of course, one believes that a full-bore defense of the classical Indian karma-rebirth ideology must be part of an account of non-self as illusion. (Physicalism does not by itself make karma and rebirth impossible, but it does make it seem implausible.) Unless one thinks there can be no Buddhism without the karma-rebirth complex, it is not clear why a Buddhist physicalism is necessarily ruled out. The Buddha was, after all, committed to the existence of five kinds of psychophysical aggregate. Yogācāra rejects the existence of one of the five (rūpa), yet no one denies that Yogācāra is Buddhist. True, all Indian Buddhist schools were anti-physicalist. But when was it decided that the history of Buddhist philosophy ended in the 12th century?⁵

³ See Siderits 2018. For an attempt to give a positive account while avoiding the paradox of liberation, see Wagner 2018.
why the Buddha claimed that the goal of his Path is merely the cessation of suffering and not some positive good (see e.g., M II.32-5 (M79, Cūla Sakuludāyi Sutta)). Even if, as I suspect, the state of “cessation with remainder” (the state of the enlightened person while still alive) can be characterized as positively valuable, there can still be good reason to refrain from characterizing it in terms that may get in the way of its attainment.

I join my fellow respondents in thanking Karsten Struhl for his stimulating paper. In prescinding from the details of Buddhological exegesis and looking at how Buddhist thought and practice might be understood when seen through the lens of some current strains of thought, it raises important questions not only about the topic of self as illusion, but also about the various ways in which we have come to think about that topic. It is, I think, good to have our methodologies interrogated from time to time.

REFERENCES

Abbreviations

BCA: The Bodhicāryāvatāra of Śāntideva with the Commentary Pañjika of Prajñākaramati (BCAP), ed. P.L. Vaidya (Dharbanga: Mithila Institute, 1960).