

Meta-Cognition Without a Cognizer: Buddhist Non-Self and Awareness of Awareness

Mark Siderits

Seoul National University / Kyoto University

***Abstract:** The doctrine of non-self is central to the theory and practice of classical Indian Buddhism. One source of recent interest in this doctrine is the widely shared sense that belief in a self may be an obstacle to naturalizing the mind and mental phenomena. While many Buddhists might reject such a project, Buddhist efforts to explain how mental processes function in the absence of an enduring subject of experience may still be worth exploring. One issue about which Buddhist philosophers had much to say is how self-knowledge or meta-cognition (our ability to cognize our own cognitions) is possible if there is no self. The example used by Indian philosophers is that of seeing blue color: typically one is able to report not just the presence of blue, but also that one sees blue. I seem to be aware not only of the things that I see and feel and think about, but also of my seeing and feeling and thinking. Descartes took this as proof of the self. The question is whether Buddhists can account for the ability given their allegiance to non-self.*

I begin by sketching the non-self doctrine as articulated in Buddhist reductionism. This will introduce the problem of self-knowledge, followed by discussion of two Buddhist responses to the problem: the claim that cognition is reflexive in nature (that cognitions cognize themselves), and a higher-order-thought account. I claim that the second response is better suited to the Buddhist project than the first. I then investigate the evidence that can be marshalled in its defense, some of which will be drawn from current work in developmental psychology and philosophy of mind.

1.

It is well known that Buddhists deny the existence of the self. What is not well known is that Buddhists distinguish between the concept of the self and the concept of the person, and that most Buddhists take not an eliminativist but a reductionist stance toward the latter. According to the mainstream Buddhist view, while persons are not to be found in our ultimate ontology, the concept of the person plays a

sufficiently important place in our cognitive economy that it should be granted the second-tier status of ‘conceptual construction’. Their view is that while strictly speaking there are no persons, it is still perfectly understandable, given our interests and cognitive limitations, that the concept of a person should play a central role in the conceptual scheme used by most people most of the time. In order to see how this view arose, we need to begin by looking at the soteriological project that Buddhism shares with most other Indian philosophical schools.

Indian philosophizing grew out of the project of seeking release from *samsāra* (the round of rebirth). The basic idea behind the project is that life as ordinarily lived is unsatisfactory because goals are chosen based on false beliefs about who we truly are. A life devoted to sensual pleasure, for instance, is based on the assumption that the self is the sort of thing that can be made better or worse through the presence of pleasure or pain. What sets Buddhism apart from other Indian liberation projects is its rejection of the idea of a self as an essence or pole of identification. The crucial mistake we make is not that we identify with the wrong sort of thing, but that we identify at all. Buddhists claim that what they call the ‘I’-sense, the sense of being a persisting subject of experience and agent of action, is misleading and the source of existential suffering.

The Buddhist strategy for overcoming this mistake begins by distinguishing between two possible referents for the ‘I’ of the ‘I’-sense: a *self*, understood as the one part among all the psychophysical elements that grounds diachronic personal identity; and a *person*, understood as the whole that is composed of the many psychophysical elements. The best-known Buddhist argument for the nonexistence of a self is the argument from impermanence.¹ It uses a taxonomy of five kinds of psychophysical elements: one corporeal and four mental. The argument is simply that since none of these elements is permanent, and a self would have to be permanent, it follows that there is no self. That the self must be permanent follows from the rebirth assumption plus the claim that the self is the essence of a sentient being. If the sentient being persists through uncountably many lives, then the entity that is its essence must likewise persist.

Many Buddhist philosophers recognized that the refutation of the self left the ‘I’-sense largely unscathed.² They explained the cognitive impenetrability of the

¹ A clear statement of the argument is to be found at *Majjhima Nikāya* III, 15–20.

² Thus Candrakīrti likens someone who thinks refuting this ‘philosopher’s self’ will extirpate the ‘I’-sense to one who, knowing there are snakes living in the walls of their house, comforts themselves with the thought that there is no elephant inside (*Madhyamakāvātāra vṛtti* 6, 141).

‘I’-sense through their account of the *person*, understood as the concept of the mereological sum of causally connected sets of appropriately assembled psychophysical elements. As mereological nihilists, they naturally deny that there is any such mereological sum. But after giving their argument against mereological realism,³ they have some important things to say concerning the semantics of enumerative expressions.

One might take the lesson of Buddhist mereological nihilism to be that we should be eliminativists about persons. As the mereological sum of mereological sums—the assembled elements at one moment in causal connection with the assembled elements at the next moment, etc.—the person looks to be ripe for elimination from our ontology. But, says the Buddhist, to eliminate the person from our ontology is to eliminate what for most people grounds appropriation: identification by one set of elements with other sets occurring earlier and later in the causal series. As we have already seen, Buddhists do claim that appropriation is a mistake. Still, simply jettisoning the practice would have bad consequences, as illustrated with the examples of the pregnant woman, the student and the convicted criminal. For instance, eliminativism means the criminal would not appropriate and thus identify with those earlier parts of the causal series that committed the crime, would hence see their present punishment as undeserved, and would thus not be deterred from future crime. The concept *person* plays an important role in our cognitive economy. Yet Buddhists do claim that its use lies behind the problem of existential suffering. The solution is not to eliminate persons but to *reduce* them to causal series of psychophysical elements.

This is explicated through their claim that ‘person’ is an opaque enumerative term. By contrast, the word ‘pair’ is a *transparent* enumerative term. It would be at best a bad joke to say that my dresser drawer contains three things: two matching socks and a pair of socks. ‘Pair’, like ‘dozen’, ‘gross’, ‘multitude’, and ‘heap’, is merely a useful way to refer to a multiplicity. For creatures with our interests and cognitive limitations, it is a useful cognitive shortcut. That this semantic role is transparent is what explains the fact that use of the term is not taken to commit us to the existence of such things as pairs, heaps, or multitudes. At least such use does not generate a *serious* ontological commitment. It does, we might say, generate a *casual* ontological commitment. Buddhists express this by saying that things of this sort are conventionally, but not ultimately, real.

³ For details see Siderits 2015, 100–103.

Our use of ‘person’ is said to be like this, except for the crucial difference that the term’s being enumerative is opaque to us. This is, they think, the source of the ‘I’-sense, and so of existential suffering. Their recommended solution is to render the term transparently enumerative. Reduction of persons to psychophysical elements in causal succession is their strategy. The idea is that we can then retain the benefits conferred by this useful cognitive shortcut without paying the price of existential suffering.

I shall not go into the details of that strategy here. For our purposes it suffices to say that the proposed reduction involves a sort of dualist ontology in the reduction base. This is what is hinted at in the term ‘psychophysical elements’: the ultimate simples are sorted into two categories, the physical and the mental. The two categories are distinct insofar as only the physical elements are said to have such spatial properties as location. The dualism here is not substance dualism but trope dualism. Buddhists take their mereological nihilism to rule out not just composite physical objects, but, more generally, substances of any sort; these are understood to reduce to bundles of momentary trope occurrences. Chief among the mental tropes allowed in their ultimate ontology are momentary occurrences of consciousness, understood as bare awareness or registry. And if one were to try to analyze ‘thinking substance’ into its ultimate mental constituents, this looks like one likely candidate. It seems intuitively plausible that consciousness is qualitatively simple in character and is thus not reducible to anything both mental and yet simpler. Its inclusion does, however, raise considerable difficulties.

2.

These difficulties grow out of the problem of explaining how meta-cognition is possible. Not only are we (at least sometimes) conscious, we may also be aware of our being conscious, cognize our own cognitions.⁴ The simplest way of accounting for this fact involves positing a subject with cognizing as one of its modes.

⁴ Unlike recent discussions of self-knowledge, the classical Indian debate over how self-knowledge is possible does not start from the presupposition that a mental state is conscious only if that state is itself represented. While some Indian accounts of self-knowledge do have as a consequence that consciousness requires self-consciousness, it is understood that this result requires argument. What is agreed by all is that some cognitions are themselves cognized; this is the phenomenon that is thought to require explaining.

Buddhists reject such an account. They thus owe us an explanation of what it is that cognizes a given cognition. To say that it is that very cognition itself—that a cognition may function as both subject and object simultaneously—seems implausible. For it violates the widely accepted principle of irreflexivity: a simple entity cannot operate on itself. The illustrative examples are legion: the knife that cannot cut itself, the fingertip that cannot touch itself, and the like. In the absence of a valid counter-example, the principle appears to hold, and to rule out reflexive self-cognition as an account of self-knowledge.

Buddhists still needed an account of self-knowledge. They were committed to the claim that one attains enlightenment by coming to somehow *directly* grasp the fact that all existents are characterized by impermanence, suffering and non-self. Since this ‘all existents’ must include any cognition that grasps these characteristics in other things, it is not clear how such a cognition could be directly grasped as bearing the marks if it could not cognize itself. A variety of possible solutions was discussed. One such solution—Dignāga’s reflexivity thesis—came to be particularly influential. It is the view that every cognition cognizes not only its object but also itself. It grows out of Buddhist embrace of a representationalist view of perception, according to which the direct object of perceptual cognition is not an external object but instead the form borne by the cognition due to sensory contact with the external object. A cognition must therefore have two forms: that of its intentional object and that of itself as that which cognizes. Dignāga then argues that the possibility of meta-cognition can only be explained by supposing that these two forms are in fact one, that cognition illuminates its object by illuminating itself. To the objection that this is ruled out by the irreflexivity principle, his commentator Dharmakīrti proposes, as a counter-example to the principle, the case of the light that illuminates itself as well as other things in the room.

In propounding his reflexivity account of self-knowledge, Dignāga seems committed to what Carruthers (2011) calls the transparency thesis—the thesis that the mind has transparent access to its own states. While Carruthers does not himself accept this thesis, he says it may be something that humans at all times and in all cultures are strongly disposed to accept as true. Bogdan (2010) likewise rejects the thesis; his developmentally based account of self-knowledge claims instead that a child’s awareness of its own mental states develops through deployment of a theory the child first acquires in order to explain and predict the behavior of others. But he nicely illustrates the transparency thesis in his formulation of the objection he thinks will naturally arise in response to his own account of self-knowledge:

Doesn't one first need to illuminate one's mind and register what is there before—and in order to—recognize, classify and make use of what is there? Don't we first turn on the lights before seeing and recognizing what is in the room? And isn't consciousness this light that illuminates the mind, before anything else is figured out there? Isn't it the case that in the morning, when awakening, the brain first turns on the consciousness switch before we become aware—and in order to become aware—of what is going on around us and inside our minds? And when consciousness weakens or is switched off, so is the mind worth talking about? In short, mustn't one become introvertly conscious of one's own thoughts *before* determining their attitudinal profile? (Bogdan 2010, 70)

Dignāga would agree.

I shall return to Carruthers' and Bogdan's defense of their alternative opacity thesis, the thesis that the mind is opaque to itself and its states. First, though, I should point out that while Dignāga's reflexivity thesis was influential in Buddhist philosophical circles, it was not the consensus view. Some Buddhists accepted a higher-order perception (HOP) account that sees meta-cognition as introspection or 'looking within'.⁵ Still others propounded a sort of higher-order thought (HOT) account, according to which meta-cognition arises out of an abductive inference. The latter account requires some explanation. The classical Indian formulation of HOT was first developed by the non-Buddhist philosopher Kumāriila. He starts with a simple but compelling attack on the claim (shared by reflexivists and HOP theorists) that one can be directly aware of one's own mental states. This is not possible, since these are mental states, like those of another person. The idea here is that if we are using a single concept when we attribute mental states to ourselves and to others, then the criteria of application should be uniform across the allegedly huge gap between the first-person case and the third-person case. Since the criteria in the

⁵ Here I am using 'higher-order' with reference to theories of meta-cognition. The term is also used in connection with theories concerning what property a mental state must have to be a conscious state. Typically, higher-order accounts of state consciousness claim that a mental state is a conscious state only if it is itself represented by some distinct ('higher-order') state. The best-known alternative first-order account claims instead that a mental state is conscious just in case it has the dispositional property of making its content available in the global workspace, i.e., to mental modules for functions like memory, speech and action. The HOT account of self-knowledge I shall discuss begins from such a first-order account of state consciousness.

third-person case are behavioral (e.g., direction of another's gaze is a sign of their attending to an object in that direction), the same should hold in the first-person case as well. What behaviors might reveal that an object such as a fruit has been cognized? Action, such as reaching for the fruit, and verbal conduct, such as saying that a fruit is present. The key point here is that the mental state that is thereby cognized is a theoretical posit. Cognitions, whether another's or one's own, are never directly grasped. The abductive inference whereby cognitions are cognized is an inference to the best explanation. And the explanation in question is one that uses a theory of mind, a theory first developed in order to help explain and predict the behavior of others in a shared physical and social environment.

Those Buddhists who adopted this account of self-knowledge rejected Kumāriila's assumption that the cognitions thus inferred are states of a self. Of course, it was still open to them to hold that consciousness itself is ultimately real. A bundle theorist need not believe in an immaterial self in order to believe that cognitions have non-physical existence. And their existence might still be consistent with their never being directly grasped. But it is also open to a Buddhist who holds the opacity thesis to claim that consciousness is no more than a theoretical construct, and so is reducible to entities of some other sort entirely. This possibility was glimpsed early on by Śrīlāta, the Sautrāntika who pointed out that if there is no direct acquaintance with consciousness, its nature can only be specified in functional terms, and that functionalization is an open invitation to reduction (Dhammajoti 2007, 163–4). The conclusion to be drawn is that consciousness is not ultimately real.⁶ And it is, once again, difficult to imagine what non-physical trope consciousness might be reducible to, given its apparent simplicity in character. There is, moreover, a specifically Buddhist motivation at work here as well. Not only does Dignāga's reflexivity thesis violate the irreflexivity principle, it also threatens to reinforce the notion of an inner subjective realm accessible only through first-person modes of awareness. Some modern scholars (e.g., Dreyfus 2011) take Dignāga's claim that every cognition is reflexively self-aware to be true to the

⁶ The Buddhists who followed Kumāriila in explaining first-person attribution of cognition as abductive inference were global anti-realist Mādhyamikas. As such, they deny that anything is ultimately real, including the physical as well as the mental. They would thus find suspect the physicalism of those who wish to naturalize the mental. Both parties can, though, agree on the negative thesis that consciousness is not ultimately real. Whether a naturalizing project can withstand Madhyamaka critique of all ultimate ontologies is a question that lies beyond the scope of this paper.

phenomenology of our experiential lives. Many Buddhists, though, see it as just another reinscribing of the 'I'-sense.

Why, though, should consciousness be reduced and not eliminated? In a word, because the theory that posits conscious mental states, our 'folk psychology', is useful. Even as we have begun to discover the nature of the underlying neurophysiological states on which conscious mental states supervene, we find it difficult to imagine how knowledge of those states, considered strictly as neurophysiological states, could ever play the role that our folk psychology plays in everyday life. Our folk psychology reliably facilitates complex feats of self-regulation and social interaction. The computational demands would be overwhelming for systems like ours were we to jettison the theory and try to use a completed brain science instead. The states that folk psychology attributes to us are, then, conventionally real. What they reduce to, however, are physical states, and this probably accounts for the fact that Buddhists did not embrace this reductionist strategy. Physicalism makes the karma/rebirth ideology appear implausible, and Indian Buddhists took that ideology quite seriously. Further exploration of this way of accounting for meta-cognition requires that we turn to more recent discussions.

3.

Both Carruthers and Bogdan espouse versions of the opacity thesis that are compatible with the view of some Buddhists that consciousness is a conceptual construction. Bogdan builds his case on the basis of what we now know about human cognitive development. The opacity thesis posits an asymmetry between self-knowledge and other-knowledge: the ability to reliably attribute mental states to ourselves is built up out of tools that were first acquired for the purpose of explaining and predicting the behavior of others.⁷ Bogdan supports the claim that there is such an asymmetry by examining the record concerning cognitive development in early childhood. Classical Indian Buddhist philosophers were unaware of many of the facts that make up this record. It would clearly be a mistake

⁷ Proponents of the transparency thesis typically posit a reverse asymmetry: subjects are directly acquainted with their own conscious states, and only subsequently come to infer such states in others by analogy. The difficulties with the analogical approach are well known. These are captured in the thought that attributions of mental states must then employ two distinct concepts, one for the first-person case and another for the third-person case. This strikes many as tantamount to solipsism.

to suppose that the case Bogdan makes is what these Buddhists ‘really had in mind’. Their concern was just to develop an account of meta-cognition that was consistent with Buddhist non-self. Our question is whether there is now any evidence supporting their account.

Bogdan’s story goes roughly like this. Because of the human obstetrical dilemma—bipedalism leads to selection for greater brain size, leading to a requirement for earlier birth—human infants are born premature (relative to other primates) and are thus dependent on adult caregivers. This leads in turn to selection for innate mechanisms that promote attachment and ensure that the infant’s subsistence needs are met. These mechanisms include such things as differential attention to faces (present at birth) and imitation of facial gestures like smiling (present soon after birth).⁸ It is of course tempting to see this repertoire as part of a deliberate strategy on the infant’s part. (More about this temptation later.) But it is no more appropriate to attribute intention here than it is in the case of the car that only fails to start on rainy weekday mornings. These are indeed reflexes. Bogdan sees in them, though, the materials out of which is constructed the full-blown self-consciousness that the transparency thesis claims is innate. The key point for our purposes is that the kind of self-knowledge reflected in the ‘I’-sense is a construction: ‘self-consciousness appears to be a *by-product assembled* out of unrelated ontogenetic adaptations with their own histories of selection’ (Bogdan 2010, 165). Its being so opens the door to the possibility that the sense of an inner first-person realm is, as Buddhists claim, deceptive.

Three phenomena will help explain the opacity thesis and its counter-intuitive take on self-consciousness: infant amnesia, absentmindedness, and blindsight. The first of these should be puzzling if we take the transparency thesis at face value: if our own conscious states are transparently given to us, and experiences are conscious states, why do we lack episodic memory for experiences had prior to about age three? Other sorts of long-term memory, e.g., that involved in aversive conditioning, seem to function properly below that age. The suggestion is that

⁸ The selectionist explanation of such phenomena helps answer the objection to theory-of-mind accounts of cognition of others’ mental states raised by Krueger and Overgaard 2012, 248. In support of their claim that we are able to perceive the mental states of others (and not merely cognize them by rapidly performing an automatic inference), they point to the manifest difference between looking at a picture of a smiling face and looking at the same picture inverted. The selectionist explanation of the infant’s differential attention to even the most schematic depictions of faces helps us see why the same visual processing cannot be expected to function when the image is inverted.

full-blown autobiographical memory requires possession of a particular kind of self-concept, and this is something that is only fully developed in the child around age five.⁹ Of course we can say that infants do distinguish between self and other, and so do have some sort of a sense of self; the mastery of motor control in very early childhood depends on this. Goal-directed motor activity involves a feed-forward mechanism: the anticipated sensory input consequent on success is represented in neural form at the time of initiation of activity (the infant forms a neural image of the sensory input that would be present were the nipple in the mouth). This representation is then canceled on success; non-cancellation signals the need for further refinement of the motion already undertaken.¹⁰ The mechanism involved here clearly depends on the distinction between the feed-forward representation generated at the time of initiation and the feedback representation obtained upon completion of the action. And we might call this a distinction between self and other. So we might think that the infant does, after all, possess a rudimentary sense of self. But then we should have to attribute the same sense of self to the motion detector that turns on the outdoor lights when a raccoon enters the backyard. For precisely the same sort of feed-forward mechanism is involved there as well. An inverted form of the representation of the detector's radar readout at one moment is superimposed on the uninverted readout obtained at the next moment. If these do not cancel one another out, the lights are turned on; if they do, then nothing happens. It would clearly be a mistake to attribute a sense of self to the motion detector. This despite the fact that the mechanism depends on distinguishing between an internal state of the detector (its 'memory' of the preceding moment) and its representation of the current state of the external world.

The phenomenon of absentmindedness poses an equal challenge to the transparency thesis. Take the common instance of resolving to stop at the dry cleaner's on the drive home, only to find oneself pulling into the driveway without the cleaning and equally without any recollection of having driven straight home. One was presumably aware of the road, other cars, traffic signals and the like during the drive, and yet none of those experiences seems to have registered. Some of the oddity of the phenomenon can be dispelled if we take self-knowledge to be the result not of introspection but of an abductive inference. We are sure we were aware of the road, other cars, etc., during the drive home, despite our not recalling any of these

⁹ For some of the details of development of episodic memory in childhood see Busby Grant and Suddendorf 2005.

¹⁰ For details see Jeannerod 2006, 23–44.

experiences, because this best explains our having arrived home intact. And we take this to best explain the present state of affairs because we have seen what can happen when one performs an attention-demanding task like texting while driving. Still we will wonder how we might have experiences that we cannot recall immediately afterward, if the mind is indeed transparent to itself.

The case of absentmindedness involves performances that we take to involve consciousness. The case of blindsight involves doing things that we think require conscious awareness, but in fact cannot; the lack of awareness cannot be blamed on mere lapses in memory. The blindsight patient is someone whose visual organs are intact and functioning but who lacks conscious awareness of anything visual. The condition is typically brought about by injury to a part of the brain involved in visual processing. What is remarkable about blindsight patients is that they are nevertheless able to perform actions that require visual input, such as walking down a corridor strewn with obstacles. Indeed, after weaving around a piles of books and the like, the blindsight patient will report that they walked straight down the hall. Dual systems theory can be invoked to explain the phenomenon. What is presumably missing in the blindsight patient is System 2 (ventral) visual processing, which makes vision input globally available and thus available for speech processing: it is what enables one to report on what one sees. The patient's System 1 (dorsal) visual system is still intact, though. System 1 processes take afferent input directly to efferent output without routing through the global workspace. This may show that they do not involve consciousness of sensory stimulation. The flinch response is a clear-cut example: visual input as of some object coming at one's head triggers ducking before one is aware of seeing the object. Because the routing goes directly from sensory input to efferent output, it is faster than processing that went via the global workspace and thus resulted in possible registry as a conscious experience. The value to the organism of such a mechanism is obvious. It is, though, relatively inflexible and so prone to error: we duck in 3-D movies too.

What the case of blindsight brings out is the point that consciousness may best be thought of as the property a mental state has when it makes its information content available to other processing modules through presentation in the global workspace. This view of consciousness as global availability can likewise help explain the phenomenon of absentmindedness. Tasks that are routinized, such as driving the same route one has taken every day for years, no longer require System 2 resources, so that one's sensory and motor representations are not globally broadcast. If episodic memory is only of conscious experiences, it is then no mystery that one

does not recall the experience of driving straight home—since there was no such conscious experience. This view does, though, make being conscious an extrinsic property of mental states, one possessed in dependence on the state's relations to other entities. This in turn bolsters the claim that self-knowledge—cognition of one's own conscious mental states—comes about through an abductive inference from action and speech.

Dretske makes the following remarks concerning the consequences of this view of consciousness for the development of self-knowledge:

[H]ow do I know I have a mind? If introspection tells me only what I think and feel, not that I think and feel, how do I discover that I think and feel, that I'm not a zombie? I am tempted to reply, I learned this the same way I found out a lot of other things—from my mother. She told me. I told her what I thought and experienced, but she told me that I thought and experienced these things....Three-year-olds know, and they are able to tell you, authoritatively, what they think and see (e.g., that there are cookies in the jar, that Daddy is home, etc.), before they know, before they even understand, that this is something they think and see. Somehow they learn they can preface expressions of what they think (Daddy is home) with the words "I think," words that (somewhat magically) shelter them from certain forms of correction or criticism. Parents may not actually tell their children that they think—for the children wouldn't understand them if they did—but they do teach them things (language must be one of them) that, in the end, tell them they think. Children are, at the age of two or three, experts on what they think and feel. They have to learn—if not from their mothers, then from somebody else—that they think and feel these things. Nonhuman animals never learn these things. (Dretske 2003, 140–41)

Bogdan would agree. There is, Bogdan holds, a crucial step in the transition from reliably reporting on one's beliefs and other attitudes, to representing oneself as having beliefs and other attitudes. The latter requires mastery of reflexive meta-representation: representing what are one's representations *as* representations.¹¹ It is not clear how the child can accomplish this without using resources only available through linguistic abilities. Dretske may be right that this is something the child learns to do through learning certain locutions. But this opens

¹¹ For a useful discussion of meta-representation see Dretske 1997, 43–55.

up the possibility that the self-ascription is not reliably reporting on something that was already there before the child mastered the locution. Bogdan says this about the propositions that serve as contents of beliefs and other attitudes that are self-ascribed: ‘as contents ascribed to attitudes, propositions are *constructs* of commonsense psychology. In other words, propositions are how the contents of attitudes *look* from the vantage point of commonsense psychology’ (Bogdan 2010, 134). The same should go for the attitudes that are said to have propositions as content.

The evidence Dretske cited does not, though, fully support his claim that children of two or three are ‘experts on what they think and feel’. At least they are not reliable reporters of their own past beliefs—when those beliefs have undergone change. As Bogdan points out, when three-year-old Child₁, has witnessed another child, Child₂, first observe a toy being put inside one of two boxes, then leave the room, after which Child₁ sees the toy moved to the other box, Child₁ will say that when Child₂ returns to the room they will believe the toy is in the second box. And it is only when the child has mastered the task they have failed here, attributing false beliefs to others, that they then become able to self-attribute past false beliefs under similar circumstances. The order of mastery is important. It is the child’s mastery of a theory that attributes beliefs and other attitudes to others that comes first; only afterwards does the child learn to apply that theory to themselves. This is the essential asymmetry of the developmental formulation of the opacity thesis.

This asymmetry is important to the account of self-knowledge that is compatible with Buddhist non-self. What the developmental record reveals is that the infant begins with a set of other-directed reflexes, which serve as a foundation for the emergence of a mind-reading faculty that, when enriched by conceptual resources made available through language, enables the child to explicitly attribute intentional states to others. And only then is the child able to turn this proto-theory on themselves and self-attribute. This is how the capacity for meta-representation, crucial for the child’s ‘knowing *that* they think and feel’, emerges. Employment of this folk psychology then endows the child with capacities for self-governance and self-revision that over time foster the development of a narrative self. Use of this theory thus comes to be second nature to us. And this makes us forget that the conscious mental states we attribute to ourselves and to others are the posits of a useful theory, not things with which we are directly acquainted.

We are sure we are not zombies. But what is the evidence that assures us of this? The developmental story we have been discussing claims we rely on a theory

designed, in the first instance, to help us meet our needs by helping us explain and predict the actions of others. When we apply that theory to ourselves, we deploy a model that generates precisely the intuitions that make zombiehood inconceivable. First-person reporting of an experience or an attitude is, the model tells us, possible precisely because experiences and attitudes are real denizens of an inner realm and objects of direct acquaintance by the master of that realm. We know that false theories are sometimes useful. Celestial navigation, which is fairly reliable, deploys the model of Ptolemaic astronomy. Still we tend to find it inconceivable that folk psychology could be no more than a useful model. Why is this, though? Bogdan suggests it is because we are being asked ‘to *conceive*, from felt experiences, of phenomenal consciousness as fully reducible, without residue, to physical matter and its functional arrangements. Since the challenge cannot be met, the conclusion is that phenomenal consciousness cannot be explained in such terms, which are the terms of science’ (Bogdan 2010, 170). But of course, if we use the model to test the hypothesis that we might be zombies, the outcome is guaranteed. This is hardly a fair test.

Here is one last piece of evidence for the possibility that folk psychology is a useful but nonetheless deceptive theory. When infants interact with their adult caregivers, we often treat the infant’s behavior as indicating intention. And we talk to infants, and take their babbling and their later attempts at verbal imitation, as efforts to communicate. It is difficult to resist viewing the infant as having a rich subjective realm that is still hidden to us but transparent to the infant. But it is worth considering the following possibility: that our intuitions here are just the ones we would expect if the process of turning infants into persons employed the powerful mechanism of the self-fulfilling prophecy. We do tend to become what we are taken by others to be. We know, for instance, that acquisition of its first language requires that the child be spoken to. What better way to facilitate the process than to inculcate the belief that the child already possesses a rich inner life and simply lacks the means to communicate it?

4.

There are two possible ways for Buddhists to account for meta-cognition in the absence of a self: claim that cognitions are reflexive, and that the noetic and

noematic poles of a cognition are somehow strictly identical,¹² or claim that such self-knowledge is the result of an abductive inference rendered rapid and automatic due to our routine use of the model of the mind captured in folk psychology. Those who champion the second option must somehow face down withering stares of incredulity. But the first has its own challenges. There is no non-controversial counter-example to the principle of irreflexivity; and it must somehow be explained how the distinction between a cognition's noematic and noetic poles could be illusory.¹³ These are formidable hurdles, and the Buddhists who propounded the first solution to the problem of self-knowledge struggled to overcome them. Less effort was expended on trying to make the second option more plausible. I would suggest, though, that recent work in developmental psychology and philosophy of mind shows how one could begin to counter the incredulous stare. Buddhists might want to reconsider their opposition to a naturalistic project.

References

- Bogdan, Radu (2010) *Our Own Minds: sociocultural grounds for self-consciousness*. (Cambridge, MA: MIT Press).
- Busby Grant, Janie & Suddendorf, Thomas (2005) Recalling yesterday and predicting tomorrow. *Cognitive Development*, 20.2, 362–372.
- Carruthers, Peter (2011) *The Opacity of Mind*. (Oxford: Oxford University Press).
- Dhammajoti, K.L (2007) *Abhidharma Doctrines and Controversies on Perception*. (Hong Kong: The Buddha-Dharma Centre of Hong Kong).
- Dretske, Fred (1997) *Naturalizing the Mind*. (Cambridge, MA: MIT Press).
- Dretske, Fred (2003) Externalism and Self-Knowledge, in Susan Nuccetelli (ed.), *New Essays on Semantic Externalism and Self-Knowledge*, (Cambridge, MA: MIT Press), 131–42.

¹² The identity of noetic and noematic poles, of 'grasped' (*grāhya*) and 'grasper' (*grāhaka*), is demanded by the claim that cognitions are ultimately real and irreducible, plus Buddhism's mereological nihilism. For an interesting recent attempt to defend this thesis see Strawson 2008.

¹³ MacKenzie tries to solve the second problem by claiming that while cognitions are not mereologically complex, they are 'aspectually complex' (2017, 351). The difficulty with this is that, as Kumārila pointed out long ago, it is not clear how something that is in no sense composite can give rise to the attribution of different aspects. Dharmakīrti's example of the man who may equally be thought of by the boy as 'teacher' and as 'father' depends on the man's carrying out distinct functions at different times.

- Dreyfus, Georges (2011) Self and Subjectivity: A Middle Way Approach, in Mark Siderits, Dan Zahavi and Evan Thompson (eds.), *Self, No Self? Perspectives from Analytical, Phenomenological and Indian Traditions*, (Oxford: Oxford University Press), 114–55.
- Jeannerod, Marc (2006) *Motor Cognition: what actions tell the self*. (Oxford: Oxford University Press).
- Krueger, Joel and Overgaard Søren (2012) Seeing subjectivity: Defending a perceptual account of other minds. *ProtoSociology: Consciousness and Subjectivity*, 47, 239–262.
- MacKenzie, Matt (2017) Luminous Mind: Self-Luminosity and Other-Luminosity in Indian Philosophy of Mind, in Joerg Tuske (ed.), *Indian Epistemology and Metaphysics*, (London: Bloomsbury Academic), 335–53.
- Siderits, Mark (2015) *Personal Identity and Buddhist Philosophy: Empty Persons*, 2nd edition. (Farnham, Surrey: Ashgate).
- Strawson, Galen (2008) What is the Relation Between an Experience, the Subject of the Experience, and the Content of the Experience?, in *Real Materialism and Other Essays*, (Oxford: Oxford University Press), 151–87.