

Vasubandhu on the First Person

NILANJAN DAS

Abstract

In classical South Asia, most philosophers thought that the self (if it exists at all) is what the first-person pronoun “I” stands for. It is something that persists through time, undergoes conscious thoughts and experiences, and exercises control over actions. The Buddhists accepted the “no-self” thesis: they denied that such a self is substantially real. This gave rise to a puzzle for these Buddhists. If there is nothing substantially real that “I” stands for, what are we talking about when we speak of ourselves? In this paper, I present one Buddhist answer to this question, an answer that emerges from the work of the Abhidharma thinker, Vasubandhu (4th-5th century CE).

1. Introduction

In Sanskrit philosophy, there is a debate between Buddhists and Brahmanical philosophers on the existence and the nature of the self (*ātman*). On both sides of the debate, the self is quite often understood as a constituent of a person – and sometimes as a person *simpliciter* – that has three characteristics. First, we can pick it out by means of the first-person pronoun ‘I’. Second, it serves as a subject of mental states and as an agent who performs actions. Third, it persists through time. The Buddhists, without exception, defended the view that such a self isn’t a basic constituent of reality. This is:

The No Self Thesis

There is no substantially existent (*dravyasat*) or fundamentally existent (*paramārthasat*) self.

The Brahmanical thinkers – especially, those belonging to the text traditions of Nyāya and Mīmāṃsā – rejected this claim. Suppose we grant that the Buddhists are right: there is no substantially or fundamentally existent self. But, then, can the first-person pronoun ‘I’ refer to anything at all?

The question matters. If we want our ordinary self-ascriptions of mental states to be truth-apt, we should want the first-person pronoun to have a referent. If ‘I’ didn’t refer to anything, self-ascriptions of mental states, like ‘I am in pain’, would simply be without a truth value – in the same way as ‘Santa Claus lives in the North Pole’ is without a truth value. If we

deny that self-ascriptions of mental states are truth-apt, it becomes hard to explain how inferences from first-person ascriptions of mental states to third-person ascriptions of mental states can be valid. For example, the conclusion, ‘Nilanjan is in pain’, seems to follow from the premises, ‘I am in pain’, and ‘I am Nilanjan’. But, if the premises are not truth-apt (given that ‘I’ isn’t a referring expression), it is difficult to explain how such inferences can be necessarily truth-preserving. But if we assign a referent to ‘I’, then it seems to commit us to the existence of a self.

To answer our question from a Buddhist perspective, I shall turn to the Abhidharma Buddhist philosopher Vasubandhu (4th–5th century CE).¹ In his magisterial *Commentary on the Treasury of Abhidharma* (*Abhidharmakośabhāṣya*), Vasubandhu tried to reconcile the claim that there is no substantially or fundamentally existent self with the claim that ‘I’ is a referring expression. In this paper, I explain how he did so.

2. Vasubandhu on Existence, Truth, and Language

The *No Self Thesis* – though universally shared amongst Buddhists – is fleshed out in different ways by different factions of Buddhists, partly depending on the conception of substantial or fundamental existence that they work with. In the sixth chapter (*kośasthāna*) of the *Treasury of Abhidharma* (*Abhidharmakośa*), Vasubandhu distinguishes two modes of existence (*sattā*): fundamental (*pāramārthika*) and conventional (*sāṃvṛtika*). Roughly speaking, a fundamentally or substantially existent object is a basic constituent of reality: its nature and existence is (in some sense) independent of other entities. By contrast, a conventionally existent object isn’t fundamentally existent, but rather something whose existence we accept because it is practically useful for us to do so. Here, I spell out this distinction, and explore what this view entails about the truth-aptness of our talk about ordinary objects.

In v. 6.4 of the *Treasury of Abhidharma* (*Abhidharmakośa*), Vasubandhu fleshes out the distinction between the two modes of existence as follows: ‘That object, which is such that there is no awareness of it when it is divided or when there is an exclusion of other entities [from it] by means of awareness, is conventionally existent (*sāṃvṛtisat*). For example, a pot

¹ The only other writer to have written explicitly about this topic with reference to Vasubandhu is Ganeri (2010, 2012, ch. 8). While Ganeri focuses on the Yogācāra writings of Vasubandhu, I only address the *Commentary on the Treasury of Abhidharma* (*Abhidharmakośabhāṣya*). On the controversy concerning the question whether the author of the *Commentary* is the same person as the author of these Yogācāra writings, see Gold (2014).

[and water]. What is fundamentally existent is different' (AK 6.4).² On Vasubandhu's view, a conventionally existent object is something that ceases to be an object of awareness (*buddhi*),³ i.e. an object of an experience or a thought, either (a) when it is divided or (b) when we mentally exclude other entities from it.

Vasubandhu unpacks the idea in his commentary (AKBh 334.4–6). For Vasubandhu, any fundamentally existent object (*paramārthasat*) is independent in two ways. First, it is *mereologically independent* of other things. So, it cannot be destroyed in virtue of being physically divided into parts. But an ordinary object like a pot depends on its parts for its existence. If I physically break the pot into the two pot-halves that it is made of, it will no longer exist. Then, we cannot correctly apply the concept 'pot' to the pot-halves that are left over: for example, we will no longer look at the pot-halves and think, 'That is a pot'. That is the sense in which the awareness of a pot ceases when it is physically broken down into parts. Second, any fundamentally existent object is *conceptually independent* of other things. Take water. We can mentally analyse the nature of water into other things: qualities such as visible form (*rūpa*), liquidity (*sneha*), and so on. Once we have mentally excluded these qualities from something, nothing remains of the intrinsic nature of the water. We cannot imagine something that is devoid of the characteristic qualities of water and still think of it as water. That is the sense in which the awareness of water ceases when the characteristic qualities of water, e.g., its visible form and so on, are mentally excluded from it.

² The square brackets mark an emendation to the Patna edition of the text (edited by Prahlad Pradhan): the text reads 'ghaṭārthavat' instead of 'ghaṭāmbuvat'. For the corresponding emendation in the Varanasi edition (edited by Dvārikādāsa Śāstrī), see AKBhd 889.12.

³ Here, I shall translate the Sanskrit term '*jñāna*' and its synonyms such as '*buddhi*', '*pratīti*', and '*pratyaya*' as either 'awareness' or 'state of awareness'. Typically, this Sanskrit word picks out contentful occurrent mental states, like experiences and thoughts. But, often, it is translated as 'cognition' rather than 'awareness'. This is slightly misleading because in contemporary philosophy and cognitive science, the term 'cognition' is reserved for mental states, like beliefs and judgements, whose contents can be verbally reported and directly used for reasoning and the control of action. But, for at least some Sanskrit philosophers, a *jñāna* needn't be like this: non-conceptual perceptual experiences count as *jñānas*, but do not have contents that can be verbally reported or directly used for reasoning and the control of action. So, I choose the more neutral terms 'awareness' and 'state of awareness' to refer to all *jñānas*.

For Abhidharma Buddhists like Vasubandhu, the only fundamental constituents of reality are certain entities – called the *dharmas* – that are both mereologically and conceptually independent of other things. Typically, these Ābhidharmikas divide the *dharmas* into five kinds: (a) matter (*rūpa*), (b) mental events (*citta*), (c) associated mental factors (*caitta*), (d) conditioned factors that are dissociated from the mental (*cittaviprayuktasamskāra*), and (e) causally unconditioned *dharmas*. However, amongst the Ābhidharmikas, Sautrāntika philosophers like Vasubandhu think that categories (d) and (e) are empty. For instance, Vasubandhu explicitly argues that all *dharmas* are causally conditioned and momentary: they arise from a collection of causal conditions and cannot persist through time (AKBh 92.4–5 ad AK 2.55, AKBh 193.2–194.14 ad AK 4.2–3). What, then, are these causally conditioned, momentary *dharmas* according to Vasubandhu? According to one natural interpretation, such *dharmas* are *tropes* or *property-particulars*.⁴ Vasubandhu takes matter to be of two kinds: (a) the *material elements* (*mahābhūta*) like the earth element, the water element, and so on, and (b) *derived matter* (*upādāyarūpa* or *bhautika*), i.e. physical qualities that are causally dependent on the material elements. The material elements are simply property-like entities: the earth element is solidity, the water element is fluidity, the fire element is heat, and the wind element is motion. Similarly, derived matter includes perceptible qualities like visible form (*rūpa*), odour (*gandha*), and flavour (*rasa*), which are derived from the material elements. By contrast, mental events are conscious or unconscious mental tropes, while the associated mental factors are those mental tropes that invariably accompany mental events, i.e. concepts or mental labels (*saṃjñā*), hedonic states (*vedanā*) like pleasure and pain, and conditioned forces (*saṃskāra*) like intention, mental habits, and so on.

Vasubandhu claims that all these fundamentally existent objects are mereologically and conceptually independent of other things (AKBh 334.7–11). Even after a visible material object has been whittled down to its mereologically simple constituents, or other entities like flavour have been mentally excluded from its nature, the awareness of the intrinsic nature of visible form doesn't cease. This is because the existence and the intrinsic nature of visible form are neither mereologically reducible to simpler parts nor mentally analysable into other things. Since visible form is both mereologically and conceptually independent in this way, it can be treated as fundamentally existent. The same, according to Vasubandhu, is true of other physical and mental tropes that he regards as the fundamental constituents of reality.

⁴ See Ganeri (2001), Goodman (2004), and Siderits (2007).

For Vasubandhu, any object whose existence is practically useful for us to accept but fails to be either mereologically or conceptually independent in this way, can at best be *conventionally existent*. Most objects that we ordinarily treat as existent – tables and chairs, pots and cloths, chariots and carts – are like this. However, Vasubandhu emphasises that this distinction between the two modes of existence does not mean that we speak *falsely* when we assert the existence of conventionally existent objects like a pot or water. He says:

However, the label ‘conventional’ is applied to those objects [such as a pot and water] alone. So, people who say on the basis of convention, ‘The pot and the water exist’, have in fact spoken the truth, not something false. Thus, this is conventional truth.
(AKBh 334.6–7)

On one way of understanding the point, corresponding to the two modes of existence, there are two notions of truth. The content of a state of awareness (or a linguistic utterance) is conventionally true (*samvṛtisatya*) just in case that content presupposes or entails the existence of fundamentally non-existent objects, but still is practically useful for us to accept because that content indirectly reflects or tracks the way fundamentally existent objects are. By contrast, the content of a state of awareness is fundamentally true (*paramārthasatya*) just in case that content does not presuppose or entail the existence of any fundamentally non-existent objects *and* accurately reflects how things are with the fundamentally existent objects. The contents of our ordinary states of awareness – our experiences and thoughts about ordinary objects like a pot or water – can only be conventionally true. The states of awareness that represent what is ultimately true are much harder to attain.

After introducing the distinction between the modes of existence, Vasubandhu mentions a view of earlier Buddhist teachers, according to which what is ultimately true can be grasped only by means of two kinds of awareness: an *extraordinary* state of awareness (*lokottarajñāna*) and an *ordinary* state of awareness that is obtained on the basis of that extraordinary awareness (*tatpr̥ṣṭhalabdhajñāna*) (AKBh 334.11–13). The first is an epistemically direct (i.e. non-inferential), non-conceptual insight into reality, which arises when a Buddhist practitioner achieves meditative equipoise on the path of insight (*darśanamārga*) after having listened to Buddhist teachings and rationally reflected and meditated on them. By contrast, the second kind of awareness is an ordinary experience or thought that is influenced by that earlier direct and non-conceptual awareness. In the second kind of awareness, ordinary objects like tables and chairs might appear as real, but the practitioner will no longer assent to that false appearance.

There is an important lesson to be drawn from this discussion about how Vasubandhu understands the relationship between language and reality. For Vasubandhu, the realm of fundamentally existent objects is inaccessible to language. Contents that are conveyed by means of linguistic utterances proliferate distinctions – the distinction between qualities and substances, between properties and their bearers, between actions and their agents, and so on – which do not exist amongst fundamentally existent objects. Yet, this does not mean that language is inherently deceptive. Vasubandhu makes this point while discussing the Buddhist doctrine of dependent arising. Roughly speaking, this doctrine says that everything is causally conditioned: existent objects arise by depending on other existent objects. The Sanskrit term for ‘dependent arising’ is ‘*pratīyasamutpāda*’: ‘*pratīya*’ is a gerund that means ‘having depended’ whereas ‘*samutpāda*’ just means ‘arising’. So, the term literally means ‘arising having depended on something else’. But, as a grammarian opponent notes, this implies that the object which arises and therefore serves as the agent of the action or event (*kriyā*) of arising, does so by initially entering into a state of depending on something else. This is incoherent: since that object could not have existed before it arose, it could not have antecedently depended on something else. While commenting on the grammarians’ distinction between an action (*kriyā*) (conveyed by the verb in a sentence) and the agent (*kartṛ*) (conveyed by the expression declined in the nominative case), Vasubandhu says:

Moreover, the following distinction between the agent and the action, which belongs to the grammarians (*sābdikī*), is unestablished: that which arises is the agent, and the arising is the event. And, here, we do not find any event of arising which is distinct from the object that arises. Therefore, there is no deception in linguistic usage. But, in the fact, this is the meaning of the utterance. ‘When this is present, the arising of this occurs; due to the arising of this, that arises’. This content is what is said to be ‘dependent arising’. (AKBh 138 15–17 ad AK 3.28ab)

Even though the literal content of a linguistic utterance may be misleading insofar as it involves distinctions that are fundamentally non-existent, nevertheless, in most cases of successful linguistic usage, it is possible to recover a content that isn’t misleading to the same degree: a content that fits the way things are at the level of fundamental reality. Vasubandhu would agree with other Buddhists that the relationship between language and reality is dysfunctional: language, without some independent epistemic work on our part, cannot help us see how the world really is. But it is not so dysfunctional that we cannot recover any

fundamentally true content at all from the literal contents of linguistic utterances. This idea – namely, that fundamental truths can somehow be salvaged from misleading literal contents conveyed by language – will be useful for understanding Vasubandhu's view about the first-person pronoun 'I'.

3. Reductionism about 'Self'

It's now time to return to *The No Self Thesis*. In the ninth chapter of his *Commentary* – sometimes called 'The Ascertainment of the Person' (*Pudgalaviniścaya*)⁵ – Vasubandhu argues that a person is nothing but a causally connected stream of momentary physical and mental tropes called the aggregates (*skandha*). There are five aggregates: matter (*rūpa*) and mental events (*citta*) along with associated mental factors such as hedonic states (*vedanā*), mental labels (*saṃjñā*), and conditioned forces (*samskāra*) like intentions and mental habits. These correspond to parts of the body as well as the physical and mental events that accompany it. A causally connected stream of these tropes is mereologically dependent on its constituents, and it is temporally extended. So, on Vasubandhu's view, it cannot be fundamentally existent. 'Therefore', Vasubandhu concludes, 'a person exists as a matter of conventional designation (*prajñaptisat*) just like a heap or a stream'. (AKBh 467.12–13) Where does this leave us with regard to the self?

Vasubandhu wants to reject two distinct views about the expression 'self'. Call the first *substantialism*, the view that the expression 'self' refers to a substantially or fundamentally existent object. Call the second *eliminativism*, the view that the expression 'self' refers neither to any substantially or fundamentally existent object nor to any collection of substantially or fundamentally existent objects. The first view was defended by the Brahmanical thinkers, such as the Nyāya and Vaiśeṣika philosophers, who took the self to be a fundamentally real inner constituent of a person, which serves as the subject of mental states and as an agent of actions. By contrast, the second view was defended by other Buddhists – Mādhyamikas like Nāgārjuna (2nd century CE) and Candrakīrti (6th century CE) – who argued that the self is neither any fundamentally existent object that is distinct from aggregates, nor reducible to either the aggregates or a collection of aggregates. Vasubandhu wants to defend:

Reductionism about 'Self'

⁵ For translations of this section, see Duerlinger (2003) and Kapstein (2003).

The expression ‘self’ does not refer to any substantially or fundamentally existent object, but does refer to a collection of substantially or fundamentally existent objects.

In a remarkable passage in the third chapter of his *Commentary*, Vasubandhu distinguishes his position from that of the Brahmanical thinkers (while implicitly distancing himself from other Buddhists):

Now, here, the outsiders [i.e. the non-Buddhists], having accepted the theory of the self, come forward: ‘If it is asserted that a sentient being passes to another world, then a self is established’. This very claim is refuted:

AK 3.18a. The self doesn't exist.

What sort of self [doesn't exist]? That which is imagined to make a connection with other aggregates after having thrown away these aggregates. Such an inner agent of action (*antaryāpārapuruṣa*) doesn't exist. This has been said by the Blessed One, ‘There is action, and there is maturation [of the fruits of action]. But no agent is apprehended, who throws away these aggregates and makes a connection with other aggregates elsewhere. For there is a formula pertaining to the *dharma*s (*dharmaśamketa*). [To explain:] with respect to that [rebirth], there is this formula pertaining to the *dharma*s: namely, ‘When this is present, that arises’, which is just an expanded statement of dependent arising’. If this is so, what sort of self is not refuted?

AK 3.18b. Just the mere aggregates alone.

If the mere aggregates are figuratively described (*upacaryate*) as ‘the self,’ then that is not refuted. (AKBh 129.5–14 ad AK 3.18ab)

Under the first conception, the self is an inner agent of activity (*antaryāpārapuruṣa*), which exists independently of the aggregates and acquires new aggregates when it is reborn and throws away the old aggregates when it dies. But there is a second conception of the self on which there is no such inner agent of activity. On this view, the term ‘self’ is just figuratively or non-literally applied to the aggregates themselves. Vasubandhu emphasizes that the momentary aggregates cannot literally transmigrate (AKBh 129.14ff ad AK 3.18b). But we

can figuratively say that a stream of aggregates, conditioned by afflictions, enters a womb, just as we can figuratively say that a lamp moves to a different place.

Elsewhere, Vasubandhu explains what this misleading view about the self (*ātmadrsti*) is:

The self-view ascribes unreal selfhood to real objects – aggregates such as matter [i.e., the body] and so on – treating them as an agent, as a subject, and as being under one's control. And extreme views (*antagrāhadṛṣṭi*) [e.g., the view that the self is eternal or that the self is destroyed at death] and the rest arise from that self-view. So, they are said to be lacking any corresponding real object (*avastukāḥ*). (AKBh 375.2–4 ad AK 6.58ab)

On Vasubandhu's view, there is in fact no fundamentally existent constituent of a person that can serve as a subject of mental states at different times, or as an agent of actions, or as something over which we have perfect control. So, when we talk as if there is such a self, we must be ascribing these uninstantiated properties associated with the self to the impermanent aggregates themselves. In doing so, we are treating the aggregates as what they are not: as a subject, as an agent, as something under our control. This view that there is a substantial self (*satkāyadrsti*) forms the basis of different extreme philosophical views about the self like eternalism (*sāśvatadrsti*), i.e, the Brahmanical view that the self is a permanent substantial entity that can survive death and be reborn, and annihilationism (*ucchedavāda*), i.e., the Cārvāka or Lokāyata view that the self is just the living body and therefore endures through our lifetime but then is destroyed at death. And these extreme views about the self, in turn, give rise to certain afflictions (*kleśa*), e.g., one's love for oneself and what belongs to oneself (*ātmātmīyasneha*), that ultimately are the cause of all our physical and mental suffering. According to Vasubandhu, since these afflictions are based on an illusory awareness of the aggregates as what they are not, they can be abandoned through insight (*darśanaheya*), i.e., by seeing the aggregates as what they are.

The Buddhist view that Vasubandhu prefers is the source of that insight. On this conception of the self, the term 'self' refers to a causally connected stream of aggregates. It steers clear of both the extremes of eternalism and annihilationism. Since the stream of aggregates consists of aggregates that are destroyed at every moment, there is no permanent fundamentally existent self. Yet, since a stream of aggregates can survive death and be reborn in virtue of new causal connections with new sets of aggregates, the self isn't destroyed at death.

Why should we accept this reductionist view of the self? Vasubandhu offers the following argument:

[The Brahmanical opponent:] Still, how is this known, namely that the expression ‘self’ applies to a stream of aggregates, not to any other referent?
[Reply:] On the basis of the absence of direct awareness (*pratyakṣa*) and inference.
(AKBh 461.5–6)

The argument here should be construed as an argument from the absence of evidence.

Non-Vacuity

The expression ‘self’ refers to something, i.e., an entity that has the three characteristics associated with our ordinary conception of the self: (i) being the referent of ‘I’, (ii) subjecthood and agency, and (iii) persistence through time.

Knowability

If the expression ‘self’ refers to anything, then its referent can be established either by means of direct apprehension (or perception) or inference.

Absence of Evidence

Neither direct awareness nor inference can establish that there is a constituent of a person, which exists separately from the aggregates and plays the roles assigned to the referent of ‘self’.

Role Sufficiency

The roles assigned to the referent of ‘self’ can be played by a stream of aggregates.

Conclusion

So, the expression ‘self’ refers to a stream of aggregates.

For now, let’s assume that *Non-Vacuity* is true: the term ‘self’ (as we ordinarily use it) isn’t empty, i.e., it does refer to something. What about *Knowability*? Vasubandhu assumes that, if the term ‘self’ refers to anything, its referent must be something that we can know either through direct apprehension, i.e., some form of perception, or by means of inference. To

motivate this premise, Vasubandhu seems to appeal to the principle that any *dharma* that exists can be known (or knowable) by either direct apprehension (in the absence of obstructions) or inference. This principle – as stated – seems questionable. Why couldn't there be existent *dharmas* that can neither be directly apprehended nor known by means of inference? A better version of *Knowability* would be weaker; it would say that we don't have any reason to posit a *dharma* that cannot be known either by direct apprehension or by inference. That follows from a principle of parsimony: we should not populate our ontology with entities that we have no reason to posit.

Turn now to *Absence of Evidence*. Under normal conditions, the six intentional objects of the senses – the sensible qualities such as visible form – and our own conscious mental events can be directly apprehended when they aren't obstructed from us. In the case of unobservable objects such as the senses themselves, we can make a causal inference from observed phenomena. For instance, from the absence of perception in blind or deaf people and from the presence of perception in non-blind or non-deaf people, we can infer that perception requires a distinct cause, i.e., a sense, which is missing in blind or deaf people. But a self that is distinct from the five aggregates cannot be directly inferred or inferred in this way. The thought may be spelled out as follows. If we take a human body and pull it apart, we don't perceive any such thing as a self other than the physical elements that constitute it; similarly, when we introspect, we don't come across anything other than the mental factors such as hedonic states, mental labels, etc. Similarly, as Vasubandhu claims later, there are no observed phenomena that can only be causally explained by positing the existence of a self that is distinct from the five aggregates. This supports *Absence of Evidence*.

Together, these three premises yield the conclusion that the term 'self' doesn't refer to anything other than the aggregates themselves. It doesn't yield the conclusion that Vasubandhu wants: the claim that the term 'self' refers to a stream of aggregates. This is where the final premise – *Role Sufficiency* – becomes relevant (though it is left unstated by Vasubandhu). If we could show that a stream of aggregates can perform all or many of the roles – such as being the referent of the first person pronoun 'I', being a subject of mental states, being an agent of actions – that are associated with the referent of 'self', a stream of aggregates will indeed be one of the most natural candidates for being the referent of the term 'self'. But, then, given that the term 'self' cannot refer to anything other than the aggregates themselves, we will end up with the conclusion that it must refer to a stream of aggregates. This is precisely what *Reductionism about 'Self'* says.

This, then, could be used to derive the conclusion that Vasubandhu ends the passage with: namely, that there is no self. On the one hand, if we are working with the Brahmanical conception of the self, then there cannot be a conventionally or fundamentally existent object that answers to that conception of the self, since the term ‘self’ simply doesn’t refer to any inner agent of activity that is distinct from the aggregates or from a stream of aggregates. On the other hand, if we are working with the thinner Abhidharma conception of the self, then there can be a conventionally existent object that answers to that conception of the self, since the term ‘self’ does refer to a stream of aggregates that has conventional existence. But the self still cannot be treated as fundamentally existent, because that stream of aggregates will be either mereologically dependent on its constituents, or temporally extended. In either case, there will be no fundamentally existent self.

Suppose we accept *Knowability* and *Absence of Evidence*. What we have not seen so far is any justification for either *Non-Vacuity*, i.e., the claim that the term ‘self’ refers to something that has three characteristics of the self – namely, being the referent of ‘I’, subjecthood and agency, and persistence through time – or *Role Sufficiency*, i.e., the claim that these three roles can be played by a stream of aggregates. To motivate *Non-Vacuity*, Vasubandhu would need an independent theory of linguistic reference, which explains why – despite our false beliefs about the self – our uses of the term ‘self’ do pick out something that plays these roles. To motivate *Role Sufficiency*, Vasubandhu would have to argue that the aggregates can indeed play these roles. In the next two sections, I consider whether he can show this.

4. The Reductionist Semantics of the First-Person Pronoun

Role Sufficiency commits Vasubandhu to a reductionist view about the referent of ‘I’. If a stream of aggregates is sufficient to play the roles associated with the self, and one of those roles is being the referent of ‘I’, then ‘I’ must refer to a stream of aggregates. This yields a kind of *reductionism* about ‘I’, the view that, in any context of utterance, ‘I’ does not refer to a substantially or fundamentally existent object, but does refer to a collection of substantially or fundamentally existent objects. In the ninth chapter of his *Commentary on the Treasury of Abhidharma*, Vasubandhu offers two arguments for this view.

Vasubandhu’s arguments occur in the context of a response to an objection raised by a Brahmanical thinker.

[The Brahmanical opponent:] If the self doesn’t exist, for the sake of what does one commence actions?

[Vasubandhu:] For the sake of a purpose like, ‘I would be happy’, and ‘I wouldn’t suffer’.

[The Brahmanical opponent:] What is this thing called ‘I’, which is the intentional object of the ‘I’-awareness (*ahankāra*)?

[Vasubandhu:] That awareness has the aggregates its intentional object. (AKBh 476.4–5)

The Brahmanical thinker asks how one could even undertake actions (intentionally) if there were no fundamentally existent self. Consider our self-interested actions, actions we undertake out of future-directed self-concern. When I walk out of home on a rainy day, I typically take my umbrella with me, so that I don’t get wet later. Or, when I save a part of my salary each month, I do it precisely because I wish to be comfortable when I retire. As Vasubandhu’s Brahmanical opponent notes, this kind of future-directed self-concern is driven by first-personal thoughts that take the form, ‘If I were to save money now, I would be happy later’, or ‘If I were to take my umbrella, I wouldn’t catch a cold from the rain and suffer later’. These judgements about the consequences of my actions could be accurate only if the two occurrences of ‘I’ – in the antecedent and the consequent of the conditional – were to refer to a self that persists through time, performs actions intentionally, and later experiences the happiness or the suffering that results from those actions. But, if there were no self, then how could we account for the contentfulness of such judgements? And, if those judgements weren’t contentful, then how can we rationally plan and (in some cases, successfully) undertake actions on the basis of such judgements?

Vasubandhu’s response to this challenge is to say that, even if there were no fundamentally existent self, the first-personal judgements of the form, ‘I would be happy’, or ‘I wouldn’t suffer’, needn’t be false or without a truth value. For the first-personal ingredient of those judgements – the ‘I’-awareness – could still be directed at the aggregates. In other words, when we use the concept or expression ‘I’ in our thought or discourse, what we pick out are the aggregates. Vasubandhu offers two arguments for the claim that the first-person pronoun ‘I’ refers to the aggregates.

[The Brahmanical opponent:] How is this known?

[Reply:] Because there is love (*sneha*) for them, and because there is co-referentiality [of the ‘I’-awareness] with states of awareness that take the form ‘fair’, and so on. In states of awareness that take the form, ‘I am fair’, ‘I am dark’, ‘I am fat’, ‘I am thin’, ‘I

am old', and 'I am young', this 'I'-awareness is observed to be co-referential with states of awareness that take the form, 'fair', and the like. But these are not features of a self. On that basis, too, it is known that this 'I'-awareness arises with respect to the aggregates. (AKBh 476.5–8)

The first of these arguments – which Vasubandhu does not elaborate on – is what we may call the *argument from self-love*. The rough thought is that our peculiar love or affection for ourselves isn't directed at anything that is distinct from the aggregates, but rather at the aggregates themselves. When I take an umbrella with me, I do so out of concern for my future aggregates, so that this stream of aggregates won't include states of pain in the future. Similarly, when I save money, I do so out of concern for my future aggregates, so that those aggregates would be accompanied by states of pleasure arising from comfort in the future. But, presumably, whatever 'I' refers to is the object of this kind of love or affection for oneself. This, in turn, implies that the aggregates themselves – or a stream of those aggregates – is what 'I' refers to.

Turn now to Vasubandhu's second argument: it is what we may call *an argument from self-ascriptions*. It proceeds from the observation that 'I'-awareness and states of awareness that take the form 'fair' can be co-referential (*samānādhikarana*), i.e., can be directed at the same intentional objects. If I ascribe physical properties to myself by means of ascriptions like, 'I am fat', or 'I am dark', the predicates 'am dark' and 'am fat' apply to the same object that 'I' refers to. But these properties that I am ascribing cannot be the properties of the self as conceived by the Brahmanical thinkers, a self that is distinct from the physical aggregates. So, if these ascriptions are indeed correct, the referent of 'I' cannot be distinct from the physical aggregates.

I don't find myself persuaded by these arguments, and neither should you. Start with the argument from self-love. We can reconstruct it as follows.

- P1. The referent of 'I' serves as the object of a peculiar kind of affection or concern, what Vasubandhu and other Buddhist thinkers call self-love (*ātmasneha*).
- P2. The aggregates are the object of such self-love or self-concern.
- C. So, the aggregates are the referent of 'I'.

Here, the Brahmanical thinkers wouldn't deny P1: typically Nyāya and Mīmāṃsā philosophers are happy to grant that the object of self-love is in fact what 'I' refers to.

However, such Brahmanical thinkers would resist P2. They could argue that if there is any object of self-love – more specifically, the kind of future-directed self-concern that underlies many of our ordinary actions – then it cannot be the momentary aggregates, but must be a unitary self that performs actions at an earlier time and later comes to experience the results of that action.

Vasubandhu seems to reject this line of objection elsewhere. In the third chapter of his *Commentary on the Treasury of Abhidharma*, Vasubandhu considers why *bodhisattvas* (i.e., morally ideal beings that can achieve enlightenment but delay it so as to alleviate the suffering of others) undertake altruistic actions (AKBh 182.7–15 ad AK 3.94a). His opponent here is a psychological egoist, who thinks that no one acts out of any motive other than self-concern. In response, Vasubandhu offers an account of self-interested actions. Under this account, we ordinarily undertake self-interested actions because we are unfamiliar with the true characteristics of the conditioned factors that constitute us as persons. Persons are nothing over and above streams of aggregates, and these aggregates are causally conditioned. As such, they share a number of general characteristics (*sāmānyalakṣaṇa*) such as impermanence and so on. However, we ordinarily overlook these aspects of the aggregates, and wrongly conceptualize them as a unitary self that persists through time, performs actions, and then later experiences the result of those actions. This habitual misconception of ourselves, in turn, makes us think that there genuinely is a fundamentally real distinction between ourselves and others. This belief in the self-other distinction gives rise to the peculiar kind of concern – what Vasubandhu calls self-love – towards the aggregates that we falsely take to be our unitary selves. How does such self-love manifest itself? According to Vasubandhu, it motivates us to perform prudentially rational actions, like saving money, that involve making present sacrifices for the sake of our own future well-being. However, the *bodhisattvas* are not subject to this misconception about the aggregates. Since they habituate themselves to a picture of reality on which there is no strict self-other distinction, they are able to withdraw their self-love from the aggregates and increase their concern for others. This motivates them to perform altruistic actions that require them to subject themselves to suffering for the sake of the well-being of others.

This defence of the possibility of altruism might be taken to support P2. But, on reflection, it is unsatisfactory. Suppose, for the purposes of supporting P2, we were to rely on Vasubandhu's claim that our attitude of self-love is based on a misconception of the aggregates as a unitary self. That would make the argument from self-love flagrantly circular;

for one of its premises would now be based on something very much like the conclusion of the argument. Thus, the argument strikes me as dubious.

So, consider the argument from self-ascriptions:

P1*. Our ordinary self-ascriptions – such as ‘I am fair’, ‘I am dark’, and so on – can correctly ascribe physical characteristics to the referent of ‘I.’

P2*. They can only be correct if physical aggregates can be the referent of ‘I.’

C*. So, the physical aggregates can be the referent of ‘I.’

Like the previous argument, this argument would fail to convince the Brahmanical thinkers. Vasubandhu himself anticipates two objections against it.

There is a figurative ascription of selfhood to the body, even though it [merely] assists the self. For example, [in the case of the awareness], ‘Whoever this is, it is just I, it is just this servant of mine’, there is a figurative ascription of selfhood even to someone who assists [oneself], but there is no ‘I’-awareness. And if the objective basis [i.e., the independently existing intentional object of the ‘I’-awareness] is the body, why doesn't it have some other body as its objective basis? (AKBh 476.9–476.11)

There are two distinct objections here. The first objection is an *objection from figurative use*. The claim is that ascriptions of physical properties like ‘I am fair’, and so on, are not literally true; they are figuratively true. Sometimes, we apparently ascribe to an object a property that in fact belongs to something to which the object is related. Consider the following pairs of sentences.

- (1a) I am parked out the back.
- (1b) My car is parked out the back.
- (2a) Yeats is still widely read.
- (2b) Yeats' poetry is still widely read.

In at least some contexts, (1a) can express the same meaning as (1b), and (2a) can express the same meaning as (2b). One might argue that the sentences (1a) and (2a) aren't literally true. But we can infer (given the context) that they have the same meaning as (1b) and (2b) respectively, and therefore infer that the expressions ‘I’, and ‘Yeats’ in fact are being used

non-literally or figuratively to refer to other things that are related to the semantic referents of these expressions: ‘I’ refers to the speaker’s car, while ‘Yeats’ refers to Yeats’ poetry. In the same way, as Vasubandhu notes, one can sometimes figuratively speak of one’s servant as ‘I’ because one’s servant assists oneself. The important point is that, in these cases, the speaker doesn’t literally identify Yeats with his poetry, or themselves with their car or their servant. The same diagnosis can apply to apparent self-ascriptions of physical properties. When we say, ‘I am fair’, or ‘I am dark’, we may be figuratively using ‘I’ to speak of our bodies. So, they needn’t be taken to correctly ascribe physical characteristics to the semantic referent of ‘I’. Therefore, the argument from self-ascriptions needn’t be sound.

The second objection is an *objection from arbitrariness*: if our own body could be the objective basis (*ālambana*) – the independently existing intentional object – of our ‘I’-awareness, then why could someone else’s body not also be picked out by the first-person pronoun? Here is one way of understanding the worry. On Vasubandhu’s view about the first-person pronoun, the first-person pronoun can *literally* refer to one’s body, e.g., when one says:

(3a) I am fair.

But there are other contexts where the first-person pronoun can literally refer to something that is distinct from a person’s body, e.g., when a person says:

(3b) My body is fair.

In other words, the first-person pronoun, on this view, is *doubly* context-sensitive. On the one hand, it can refer to the mental or physical aggregates associated with different persons in different contexts of utterances. On the other hand, depending on the context, it may refer to different sets of aggregates associated with the same person. This, in turn, opens up room for the following question. What are the contextual factors that determine which aggregates will be picked out by ‘I’ in a particular context? If the answer is simply that this is determined by the speaker’s intention, then the semantics of the first-person pronoun will be too unconstrained. For, now, if a speaker intends to pick someone else’s body by means the first-person pronoun ‘I’, she may indeed succeed in doing so. As a result, the pronoun may indeed end up referring to the physical aggregates associated with another person in that context. But, surely, that will be a misuse of the expression ‘I’. A natural way of blocking this result will

be to say that the first-person pronoun has a fixed referent that doesn't vary depending on the context (as Vasubandhu suggests). And the easiest way of implementing this strategy will be to say that ascriptions like (3a) are at best only figuratively or non-literally true; in such cases, the use of 'I' in such ascriptions is intended to convey the same meaning that is ordinarily conveyed by the definite description 'my body'.

It's time to take stock. In this section, we have considered two arguments for the conclusion that the first-person pronoun refers to the aggregates or a stream of aggregates. But neither of these arguments seem to succeed. There is no obvious way for Vasubandhu to defend the argument from self-love without making it flagrantly circular. And the argument from the self-ascription of physical properties can be resisted either by appealing to the possibility of figurative uses of 'I' or by appealing to the danger that, if this argument were sound, someone else's body could also be picked out by 'I'.

5. 'I' and the Causal Theory of Reference

To see how Vasubandhu might resist these objections to his two arguments, it's best to begin with his response to the second objection from arbitrariness to the argument from self-ascriptions. He writes:

[Vasubandhu:] Because there is no connection. For, whatever body or mental event with which this 'I'-awareness is connected, the 'I'-awareness arises with respect to that, and not with respect to anything else. The reason is that there has been a habituation of this kind in the beginningless cycle of rebirth.

[The Brahmanical opponent:] And what is that connection?

[Vasubandhu:] The connection between an effect and its cause. (AKBh 476.11–13)

The claim is that 'I'-awareness is directed at the physical or mental aggregates – the body or the mental events – that it is causally connected to (in the appropriate ways), but not at anything else. For Vasubandhu, the 'I'-awareness is one mental event amongst others, embedded through certain relations of causal dependence within a causally connected stream of physical and mental aggregates. So, when we think first-person thoughts, we can only pick out by 'I' other physical or mental aggregates that are part of the relevant stream, but not physical or mental aggregates that fail to overlap with that stream. That is why 'I' cannot refer to a body that doesn't overlap with the physical aggregates that are part of the stream of aggregates that the 'I'-awareness is part of. Why is that? This is because – as Vasubandhu says

– we have formed the habit of applying the concept or the conventional designation ‘I’ to the physical or mental aggregates that the relevant ‘I’-awareness is part of.

I wish to show that this theory of what ‘I’ refers to follows from a more general theory of reference that Vasubandhu subscribes to. The aim of this section is to spell out this theory and explore its consequences for the first-person pronoun.

5.1 *The Causal Theory of Reference*

We see this theory in action when Vasubandhu argues against certain Buddhists – the Vātsīputrīyas, sometimes known as the Personalists (*pudgalavādins*) – who reject the Abhidharma reductionist theory of persons.

The Vātsīputrīyas don’t think that persons just are the aggregates connected in some way. Rather, they claim that persons – even though they ontologically and conceptually depend on the aggregates – are nevertheless irreducible to the aggregates. The Vātsīputrīyas, in effect, endorse what looks like an emergentist theory of persons:⁶ persons are emergent entities that arise and are conceptualised when the aggregates enter into certain relations with each other, but nevertheless have novel properties and causal powers that cannot be explained solely in terms of the properties and the causal powers of the aggregates themselves. But, as they are represented in the ninth chapter of Vasubandhu’s *Commentary*, they state this idea in a somewhat puzzling manner. They say that the relation between the aggregates and persons is indeterminable (*anyākta*): while the conventional designation ‘person’ is applied in dependence on the aggregates, a person can be neither determinately identical (i.e., reducible) to the aggregates, nor determinately completely distinct (i.e., independent) of them.

In response to this theory, Vasubandhu asks what exactly the Vātsīputrīyas mean when they say that the conventional designation ‘person’ is applied in dependence on the aggregates (AKBh 461.20–24). Either the conventional designation ‘person’ is applied after one has apprehended the aggregates, or its application is caused by the aggregates themselves. In either case, the conventional designation will only pick out the aggregates and nothing else. The example that Vasubandhu gives of the first kind of application is the application of the conventional designation ‘milk’. On Vasubandhu’s view, we apply the conventional designation ‘milk’ in our ordinary discourse by initially becoming perceptually aware of the sensible qualities of milk such as its visible form, its taste, and so on. If that is right, then, Vasubandhu claims, our conventional designation ‘milk’ can only pick out those qualities

⁶ For discussion, see Priestley (1999), Châu and Boin-Webb (1999), and Carpenter (2015).

alone, not any substance that exists over and above them, since only the initial awareness of those qualities triggers the application of the relevant term. Similarly, if the presence of the aggregates themselves triggers the application of ‘person’ without any mediating awareness of the aggregates themselves, the term can only refer to the aggregates themselves, since they alone are responsible for the application of the term.

Vasubandhu later returns to this point, again, while responding to the Vatsīputrīyas’ claim that the relationship between a person and the aggregates is indeterminable: the person is neither determinately identical to the aggregates nor determinately distinct from them (AKBh 463.10–17). But, there, Vasubandhu claims that there is no room for such indeterminacy given the following principle.

The Causal Theory of Reference

A conventional designation or a concept E refers an object of kind K if, in competent users of E , applications of E are (typically or normally) caused by an initial awareness of objects of kind K .

Take the case of milk and water. Suppose that we are competent users of the expressions ‘milk’ and ‘water’, and that we typically apply these expressions in ordinary discourse on the basis of detecting their sensible qualities like colours, tactile qualities, flavours, odours and so on. So, Vasubandhu’s causal theory of reference would predict that these expressions refer to these four kinds of sensible qualities that can be detected by the four senses, and nothing else. Moreover, we shouldn’t assume that the term ‘milk’ or ‘water’ is ambiguous: it is not the case that, in some contexts, the term ‘milk’ refers the visible form of milk; in some, it refers just to its tactile qualities; in others, just to its flavour or odour. We are not aware of any such ambiguity in our linguistic usage. Rather, the right conclusion to draw is that the term ‘milk’ simply refers to the collection of these qualities, i.e., to these sensible qualities arranged together in a certain way. Vasubandhu explicitly claims that, from this, it straightforwardly follows that milk and water just are these qualities themselves collected together in a certain way; there is simply no scope for any (epistemic or metaphysical) indeterminacy about whether milk and water are identical to the sensible qualities that can be detected by the four senses or not. The facts about our application of these expressions decisively settle the matter.

Given the *Causal Theory of Reference*, Vasubandhu would want to draw the same lesson about our use of the conventional designation ‘person’. Assume that we are competent users of the term ‘person’. Suppose also that we typically or normally apply this expression either

on the basis of our perceptual awareness of physical aggregates like the parts of our bodies arranged in a certain way, or on the basis of our introspective awareness (*manovijñāna*) of the mental aggregates such as our conscious mental occurrences. Then, it follows from Vasubandhu's theory of reference that this conventional designation should simply refer to those objects of those kinds: the aggregates collected together. If that is so, it is determinately true that a person just is the physical and mental aggregates arranged in a way, which can be either apprehended by one's senses such as sight, or by means of introspective awareness. So, there is no room for any epistemic or metaphysical indeterminacy about whether the person is identical to those aggregates or not.

A feature of the *Causal Theory of Reference* is worth highlighting. Earlier, I said that, on Vasubandhu's view, even though language may not be a reliable guide to the structure of reality, the relationship between language and reality is not so dysfunctional that our linguistic utterances cannot indirectly (or partly) reflect the way fundamentally existent objects are. The *Causal Theory of Reference* helps us to see why that might be the case. It allows Vasubandhu to do two things at once: it allows him to give a reductionist semantics for expressions like 'pot', 'chariot', and 'person', and it allows him to say that we (ordinary people as well as philosophers) can be subject to massive error about what these expressions refer to. On the one hand, the *Causal Theory of Reference* will predict that these terms pick out certain collections of mental or physical tropes insofar as they are typically applied on the basis of our initial awareness of those tropes. So, using these linguistic expressions, we can indeed talk about fundamentally existent objects. On the other hand, it is compatible with the *Causal Theory of Reference* that we have lots of false beliefs about these collections of tropes. We might think that ordinary objects like pots and chariots exist over and above their parts and persist through time. Similarly, we might think that selves and persons are persisting entities that serve as subjects of mental states and perform actions, but aren't reducible to the aggregates themselves. But these false beliefs needn't prevent expressions like 'pot', 'chariot', and 'person' from referring to collections of fundamentally existent objects; for, despite having these beliefs, we may continue to apply these terms – in ordinary discourse – on the basis of our initial awareness of the relevant collections of fundamentally existent objects. Thus, *Causal Theory of Reference* permits our uses of language to track truths about how the world fundamentally is, while leaving room for us to be massively mistaken about the structure of reality.

Notice that we can apply the *Causal Theory of Reference* to the first-person pronoun 'I'. Even though we are massively mistaken about who we are, 'I' nevertheless refers to a stream

of aggregates. ‘I’ is applied in our first-person thoughts, either on the basis of our sensory awareness of our body (e.g., when one thinks, ‘I am fair’) or on the basis of our introspective awareness of our inner conscious mental occurrences (e.g., when one thinks, ‘I am in pain’). In either case, the application of ‘I’ is typically based on certain states of awareness that are directed at aggregates belonging to a particular stream. So, ‘I’ should refer to those aggregates. This just means that, if we accept the *Causal Theory of Reference* and Vasubandhu’s general theory about how we apply the first-person pronoun, the eliminativist view about the reference of ‘I’ cannot be right. In the rest of this section, I will explain how this theory of reference allows Vasubandhu to respond to the objections to at least one of his arguments for the reductionist semantics of ‘I’.

5.2 *The Objection from Figurative Use*

The *Causal Theory of Reference* enables Vasubandhu to defend his argument from the self-ascriptions of physical properties. The initial objection to that argument was the objection from figurative use: namely, that our self-ascriptions of physical properties aren’t literally true. On this view, an ascription like ‘I am fair’, is like the statement, ‘I am parked outside’. The latter is figuratively or non-literally true insofar as it conveys the content that one’s car is parked outside. Similarly, the self-ascription of fairness is figuratively or non-literally true insofar as it conveys the content that one’s body is fair.

In response, Vasubandhu could begin by noting that ‘I’ is context-sensitive in two ways. First, ‘I’ is context-sensitive because ‘I’ can pick out aggregates associated with different person in different contexts. When ‘I’ is used by Devadatta, it will pick out the aggregates associated with Devadatta; when it is used by Yajñadatta, it will pick out the aggregates associated with Yajñadatta. But ‘I’ can be context-sensitive in a different way. In some contexts of utterance, Devadatta may correctly say, ‘I am fair’. But, yet in other contexts of utterance, the same Devadatta may correctly say, ‘My body is fair’. In contexts of the latter sort, the second ascription can only be literally correct if the referent of ‘I’ is something that distinct from Devadatta’s body. But, then, the first self-ascription cannot be correct in those contexts, since it entails the referent of ‘I’ is a bearer of Devadatta’s physical properties and therefore is Devadatta’s body. The best way to preserve the truth of both ascriptions is to say that ‘I’ is context-sensitive in a different sense. In the contexts where ‘I am fair’ is true, ‘I’ refers to the body, or a stream of aggregates that includes the body. But, in the contexts where ‘My body is fair’ is true, ‘I’ doesn’t refer to the body, or a stream of aggregates that includes the body, but rather to something that is distinct from the body.

Vasubandhu's theory of reference can explain these shifts of reference without treating any of these uses of 'I' as non-literal or figurative. Vasubandhu has already told us that 'I' can refer to physical aggregates as long as there is the right sort of causal connection between the relevant 'I'-awareness and those physical aggregates. But, in another context, Vasubandhu notes that, at least sometimes, 'I' only refers to mental events.

Since mental events are the basis (*samniśraya*) [i.e., the intentional object] of 'I'-awareness, they are figuratively described as the self. It has been said [in the *Dhammapada*]: 'For a wise person attains heaven by means of a self that is well-restrained'. And, in another context, the Blessed One has spoken of the restraint of mental events [also in the *Dhammapada*]: 'The restraint of mental events is good, and mental events, when restrained, bring happiness'. (AKBh 27.6–12 ad AK 1.39a)

This suggests that, in the final analysis, the first-person pronoun can refer to different collections of aggregates in different contexts of utterance. What a certain use of 'I' refers to depends on which aggregates serve as the proximate causes of the relevant use of the pronoun. For example, if the use of 'I' is triggered by the perceptual awareness of one's body, then it may indeed refer to an embodied entity. But, when it is triggered only by one's introspective awareness of one's conscious mental life, then it needn't refer to an embodied entity. Thus, depending on the context, 'I' may pick out mental aggregates, or physical aggregates, or both. We can state the theory as follows.

The Semantics of 'I'

In any context of thought or utterance, an application of the expression or concept 'I' refers to a contextually salient collection of (i) purely mental, or (ii) purely physical, or (iii) mental and physical aggregates, which serve (in the right way) as the proximate causes of the relevant 'I'-awareness or 'I'-thought.

Thus, Vasubandhu can block the objection from figurative use by appealing to the two kinds of context-sensitivity that affect the use of 'I'.

5.3 The Objection from Arbitrariness

Consider next the objection from arbitrariness. Given that the expression 'I' is context-sensitive on Vasubandhu's view, it can refer sometimes to both the mental and physical

aggregates that are part of the stream within which the relevant ‘I’-awareness occurs, sometimes only to the mental aggregates, and sometimes only to the physical aggregates. Then, why can’t it refer to the body associated with another stream? Vasubandhu has told us (in the passage quoted at the beginning of this section) that this is because the appropriate kind of causal connection between another person’s body and the ‘I’-awareness is absent. However, we might not find this response satisfying.

For instance, one could argue that our ‘I’-thoughts can indeed be causally connected in some way to other people’s bodies. Imagine an agent who is wired up in such a way, such that, on the basis of her perceptual awareness of other people’s bodies, she thinks thoughts like, ‘I am fair’, or ‘I am thin’. Surely, if that person’s own body is not fair or thin, we would want to say that the contents of her ‘I’-thoughts are false. This means that this person is misapplying the expression ‘I’; she is conceptually or linguistically incompetent. Yet, in this case, there is a causal connection between the other people’s bodies and one’s ‘I’-awareness: just as an ‘I’-awareness may arise for a person with respect to her own body through the mediation of the perceptual awareness of her own body, so too, in this case, an ‘I’-awareness for the deviant user of ‘I’ arises with respect to another’s body. Vasubandhu needs to say why conceptually or linguistically competent agents never use the first-person pronoun in the same way as this deviant user of ‘I’ uses it. Why are causal connections that hold between the ‘I’-awareness and the mental or physical aggregates that it picks out special in such a way that ‘I’ cannot refer to other people’s bodies?

Notice that there is no analogous challenge for the Brahmanical thinker. For a Nyāya or Vaiśeṣika thinker, any use of the concept ‘I’ in a thought refers to a subject who serves as the owner of that thought. Who is this subject? It’s just a particular self in which various mental qualities – including states of awareness – reside. The fact that the ‘I’-thought resides in that self makes that self exclusively the owner of the relevant ‘I’-thought. So, on this view, ‘I’ refers neither to one’s own body nor to someone else’s body. So, a defender of this view asks Vasubandhu: ‘If the self doesn’t exist, whose is this ‘I’-awareness?’ (AKBh 476.13–14) The implicit thought seems to be that causal connections alone aren’t enough to single out an owner of any ‘I’-thought, which the first-person pronoun ‘I’ could then refer to.

In reply, Vasubandhu appeals to a theory of ownership that he develops elsewhere while discussing an analogous challenge about the ownership of memory (*smyti*). In response to a Nyāya opponent, Vasubandhu argues that, in order to explain memory, we don’t have to appeal to a self. In the absence of impediments, e.g., disease, grief, etc., a memory arises from a mental event – e.g., an impression (*vāsanā*) – which is characterised by a mental label

(*samjñā*), etc. that resembles and is causally connected to an earlier act of attending to the relevant intentional object (AKBh 472.16–22). Suppose I tasted a raspberry sorbet. When I attended to that sorbet, I determined it to have a number of characteristics. This is what is meant by mental labelling: in determining the object to have a number of characteristics, I attached a bunch of mental labels (which are concept-like representational devices) to it. This determination may have left an impression (i.e., a memory trace) in the stream of aggregates that constitute me. That impression resembles the earlier determination I made when I tasted the sorbet and is causally connected to it. This impression, though momentary, may be replicated over and over again, and carried down my stream of aggregates. When that impression is ready to give rise to a conscious memory, then (in the absence of impediments) it will give rise to that memory. This explanation of memory doesn't appeal to any self that is distinct from the stream of aggregates.

The Nyāya opponent here notices that these relations of resemblance and causation between the impression and the earlier experiences aren't enough to explain the phenomenon of memory. Take the following case. Yajñadatta tells Devadatta about his childhood experience of blowing the candles out on his fifth birthday. As a result of that testimony, Devadatta comes to an impression that later gives rise to a non-veridical awareness where he seems to remember himself blowing out the candles on his fifth birthday. Surely, we cannot say that Devadatta remembers blowing out the candle on his fifth birthday. As Nyāya philosophers would tell us, 'One person doesn't remember what another has experienced'. The only obvious way of getting around this problem to appeal to the owner of the relevant memory: a conscious memory can only belong to someone who is originally served as the subject of the experience that the memory is derived from. Now, the Nyāya philosopher would have an easy time explaining who the owner of the memory is: it is the unitary self that serves as the substratum of the original experience as well as the later memory that arises from it. But Vasubandhu cannot appeal to such a self. So, the objector asks: 'If the self does not exist, whose is this memory?' (AKBh 473.2).

In response, Vasubandhu develops a theory that includes two claims. The first claim is that it is a mistake to construe the relation of ownership between a subject and her own mental states in the way as we understand the relation of ownership between a person and her external possessions. The argument, in a nutshell, is this (AKBh 473.2–6). The relation of ownership between a person and her ordinary possessions (like a cow) consists in her ability to use that possession for whatever purpose she wants. Since there may no purpose external to a mental state itself for which the mental state may be used, this relation of ownership doesn't

hold between subjects and mental states. So, the opponent's proposal fails. This suggests that Vasubandhu is mildly sceptical of the idea that there is anything like a relation of ownership that holds between a person and her mental states.

Vasubandhu's second claim arises from a slightly more conciliatory position. Perhaps, we can grant that a mental state is owned by a subject only insofar as it can be used by the subject (e.g., for the purposes of thinking or reasoning about a certain object). A mental state can be used in this way only insofar as the subject brings about the relevant mental state (in the right way). But, then, the ownership consists in nothing but a causal relation. This yields the second claim: the relation of ownership between a person and her mental states can be reduced to a causal relation (AKBh 473.7–10). This allows Vasubandhu to offer a reductionist account of ownership. On Vasubandhu's view, Caitra is said to be the owner of the cow insofar as the physical and mental aggregates that constitute Caitra serve as the cause of various changes in the aggregates that constitute the cow.

Furthermore, when that stream of collections of causally conditioned factors – called ‘Caitra’ – is said to be the owner of the stream called a ‘cow’, that too is done keeping in mind [Caitra's] status as a cause in relation of changes that are taking place in a different spatial region [i.e., in the cow]. However, there is no unitary object called Caitra, nor is there any cow. Therefore, even in that case, there is no status of ownership beyond the status of being a cause. (AKBh 473.10–13)

Vasubandhu will grant that streams of aggregates other than that which we call ‘Caitra’ can be causally connected to the stream of aggregates that is the cow. But that doesn't mean that those streams of aggregates can be treated as owners of the cow. The relation of ownership which holds between Caitra and the cow holds not in virtue of any *arbitrary* causal connection, but only certain *specific* causal connections. But, importantly, for Vasubandhu, this relation of ownership consists in *nothing but* causal connections. This idea, on his view, transfers over to the case of persons and their mental states. While other people can be causally connected to our memories, those memories cannot be said to belong to those other people. This is because, even though the relation of ownership that connects a memory to its owner is a causal connection, not every causal connection is sufficient for ownership. However, it still remains true that the relationship between a person and her mental states consists in *nothing but* causal connections.

Let's see how this applies to the case of 'I'-awareness. Just as the owner of a conscious memory is just a collection of aggregates that are causally related to it in specific ways, so also the owner of the 'I'-awareness is a collection of aggregates that is causally operative in the right way.

[The opponent:] And what is that cause other than [a self]?

[Reply:] A defective mental event, which is causally conditioned by earlier 'I'-awareness and has its own stream as its intentional object. (AKBh 476.15–16)

This suggests a broader story about how we come to ascribe physical or mental properties to ourselves. On Vasubandhu's view, a memory (which may or may not be first-personal in content) arises from an immediately preceding mental event, i.e., a memory trace which is causally connected to a past determination of an object that was experienced in the past and which is therefore directed at that object. Similarly, a first-personal thought that takes the form, 'I am fair', or 'I am in pain', arises from an immediately preceding mental event, i.e., a memory trace which is left by previous states of 'I'-awareness that were directed at the same stream of aggregates within which the memory trace occurs. Therefore, the memory trace is directed at the stream of aggregates within which it occurs. Since this memory trace is connected to the false conception of the aggregates as a unitary persisting self, it is said to be a 'defective mental event'.

This analogy between memories and 'I'-thoughts helps us to see how Vasubandhu might be able to respond to the objection from arbitrariness satisfactorily. On Vasubandhu's view, our competence with the concept or expression 'I' isn't something that we acquire through our exposure to the linguistic behaviour of other competent users of the relevant concept or expression. Rather, it is inherited from past lives in virtue of memory traces that are left behind by our previous 'I'-thoughts. These memory traces – as Vasubandhu notes – involve the disposition to conceptualize the momentary aggregates as a unitary self that persists through time. Thus, we are born with an innate disposition to use the word 'I' to pick out a certain stream of mental or physical aggregates while, at the same time, remaining unaware of certain key characteristics of those aggregates (such as their impermanence). This innate disposition is part of our beginningless ignorance (*avidyā*) regarding who we are: it forces upon us the view that we are unitary selves. Importantly, the memory traces – which underlie our uses of 'I' – are only directed at the stream of aggregates within it occurs, but not

to other streams of aggregates. That explains why we (as competent users of ‘I’) cannot use ‘I’ to pick other people’s bodies. So, there is no arbitrariness in the way we use ‘I’.

6. Conclusion

We are now in a position to see how this account allows Vasubandhu to reconcile the *No Self Thesis* with the view that ‘I’ is a referring expression. On this view, ‘I’ refers to a stream of aggregates. Thus, the view can explain how our ascriptions of mental states and actions to ourselves can be conventionally true (or false). Consider self-ascriptions of mental states like:

- (4) I am pleased.
- (5) I am in pain.

The Nyāya philosophers assume that (4) and (5) ascribe mental states of pleasure and pain to a substratum or basis (*āśraya*) in which they reside: a distinct substance called the self. By contrast, for Vasubandhu, these self-ascriptions are like:

- (6) The forest has borne fruit.

He explains the point as follows:

[The Brahmanical opponent:] If the self doesn’t exist, who is this being that is pleased or in pain?

[Vasubandhu:] It is that basis in which pleasure or pain arises, just as in the case of the ascriptions, ‘The tree has blossomed’, ‘The forest has borne fruit’, and so on.

[The Brahmanical thinker:] Still, what is the basis of these two mental events?

[Vasubandhu:] The six sense-bases [i.e., the senses]. (AKBh 476.16–18)

In (6), the expression ‘the forest’ refers to a collection of trees while the ascription of the predicate ‘has borne fruit’ conveys that some fruit has arisen having as its causal basis one or more of the trees in that collection. Similarly, in self-ascriptions of mental states like (4) and (5), the expression ‘I’ refers to stream of physical aggregates (e.g., sense-bases) in dependence on which these mental events of pleasure and pain can arise, while ‘am pleased’ or ‘am in pain’ conveys that one or more those aggregates have given rise to a state of pleasure or pain. But note that ascriptions like (4)–(6) cannot be ultimately true: they presuppose or entail the

existence of ultimately non-existent objects like forests and streams of aggregates. However, they can still be conventionally true or false, since they indirectly track or reflect certain facts about fundamentally real entities such as the material *dharma*s that constitute the forest, or the physical and mental aggregates that are picked out by ‘I’.

This account would allow Vasubandhu to explain how inferences of the following kind can be truth-preserving: ‘I am in pain. I am Nilanjan. Therefore, Nilanjan is in pain’. In this case, both ‘I’ and ‘Nilanjan’ are expressions that pick out streams of mental and physical aggregates. So, if a hedonic state of pain occurs in the stream of aggregates that is referred to by ‘I’, and that stream is identical to the stream of aggregates picked out by ‘Nilanjan’, then the conclusion will necessarily be true. Thus, Vasubandhu’s reductionist semantics for ‘I’ can explain the truth-aptness of our ordinary ascriptions of mental and physical properties to ourselves as well as the validity of our ordinary inferences about ourselves.⁷

References

Primary Texts and Abbreviations

- AK Vasubandhu's *Abhidharmakośakārikā* in *Abhidharmakośabhāṣyam of Vasubandhu*, edited by Prahlad Pradhan and Aruna Haldar, second edition, (Patna: K. P. Jayaswal Institute, 1975).
- AKBh Vasubandhu's *Abhidharmakośabhāṣya* in *Abhidharmakośabhāṣyam of Vasubandhu*, edited by Prahlad Pradhan and Aruna Haldar, second edition (Patna: K. P. Jayaswal Institute, 1975).
- AKBh_D Vasubandhu's *Abhidharmakośabhāṣya* in *Abhidharmakośa and Bhāṣya of Ācārya Vasubandhu with Sputārthā Commentary of Ācārya Yaśomitra*, edited by Svāmī Dvārikādāsa Śāstrī (Varanasi: Bauddha Bhāratī, 1987).

⁷ I would like to thank Joachim Aufderheide, Julian Baggini, Sophia Connell, Davey Tomlinson, Shaul Tor, and the audiences at the Ancient Philosophy Seminar in the Institute of Classical Studies, London, and at the Royal Institute of Philosophy London Lectures, for their helpful comments.

Sphu Yaśomitra's *Sputārthā* in Svāmī Dvārikādāsa Śāstri (ed.), *Abhidharmaśāstra and Bhāṣya of Ācārya Vasubandhu with Sputārthā Commentary of Ācārya Yaśomitra* (Varanasi: Bauddha Bhāratī, 1987).

Secondary Literature and Other Sources

Amber D. Carpenter, 'Persons Keeping Their Karma Together', in Koji Tanaka, Yasuo Deguchi, Jay L. Garfield, and Graham Priest (eds.), *The Moon Points Back* (New York: Oxford University Press, 2015), pp.1-44.

K. L. Dhammadhoti, *Sarvāstivāda Abhidharma*, third edition, (Hong Kong: Centre for Buddhist Studies, University of Hong Kong, 2007).

James Duerlinger, *Indian Buddhist Theories of Persons: Vasubandhu's 'Refutation of the Theory of a Self'* (London: Routledge, 2003).

Jonardon Ganeri, *Philosophy in Classical India: An Introduction and Analysis* (London: Routledge, 2001).

Jonardon Ganeri, 'Subjectivity, Selfhood and the Use of the Word 'I'', in Mark Siderits, Evan Thompson, and Dan Zahavi (eds.), *Self, No Self?: Perspectives From Analytical, Phenomenological, and Indian Traditions* (Oxford: Oxford University Press, 2010).

Jonardon Ganeri, *The Self: Naturalism, Consciousness, and the First-Person Stance* (Oxford: Oxford University Press, 2012).

Charles Goodman, 'The Treasury of Metaphysics and the Physical World', *The Philosophical Quarterly*, 54:216 (2004), pp. 389–401.

Jonathan Gold, *Paving the Great Way: Vasubandhu's Unifying Buddhist Philosophy* (New York: Columbia University Press, 2014).

Mark Siderits, *Buddhism as Philosophy: An Introduction* (London: Ashgate, 2007).

Leonard Priestley, *Pudgalavāda Buddhism: The Reality of the Indeterminate Self* (Toronto: Centre for South Asian Studies, University of Toronto, 1999).

Thích Thiện Châu and Sara Boin-Webb, *The Literature of the Personalists of Early Buddhism* (Delhi: Motilal Banarsidass, 1999).