

## Is the Buddhist Doctrine of Non-Self Conceptually Incoherent?

PAUL BERNIER

UNIVERSITÉ DE MONCTON

paul.bernier@umoncton.ca

### ABSTRACT

*Virtually all schools of Buddhism do not accept a permanent, substantial self, and see everything as non-self (anatta). In the first part of this article I recall some arguments traditionally given in support of this perspective. Descartes' cogito argument contradicts this, by suggesting that we know infallibly that the self, understood as a substantial enduring entity, does exist. The German aphorist Lichtenberg has suggested that all Descartes could claim to have established was the impersonal 'There is thinking' (Es denkt), which would support the perspective of non-self. Bernard Williams has argued that Lichtenberg's impersonal version of the cogito is conceptually incoherent, which would entail that the Buddhist perspective of non-self is also incoherent. I propose to defend the coherence of the Buddhist perspective of non-self against Williams's argument.*

### Keywords

*anatta, non-self, cogito, self, personal identity, Descartes, Lichtenberg*

According to the Buddhist perspective of non-self, a substantial enduring entity underlying one's various experiences cannot be found at all: all *dhammas*, including *nibbāna*, are non-self. This view is accepted by virtually all classical and contemporary schools of Buddhism.<sup>1</sup> In early Buddhism, as we find it expressed in the *suttas* of the Pali Canon, the systematic undermining of all claims that anything is a permanent, substantial self, or related to such a thing, is carried out prima-

---

1. There is however a notorious exception. As is well known, at an early stage of the development of Buddhism many different schools have emerged. One of these was the Pudgalavāda, or 'doctrine of persons' school, which seem to have rejected the explicit denial that there is a substantial self. But, the interpretation of this view is highly controversial, especially given the paucity of direct textual sources. Also, in some forms of East Asian Buddhism, the *Tathāgata-garbha* or Buddha-nature is talked of as one's 'true Self'.

rily in a soteriological context. That is, in order to achieve *nibbāna*, or liberation from the cycle of birth and rebirth, which is understood as painful/stressful (*dukkha*), one should directly see that there is not anything that can rightly be taken as a permanent, substantial self. Nevertheless, this claim is also relevant from a purely philosophical point of view, especially with respect to the issue of personal identity.

Descartes' famous *cogito* argument (*cogito ergo sum*: 'I think, therefore I am'), if correct, would undermine the very coherence of the Buddhist perspective of non-self. As we know, Descartes' motivation for the *cogito* was essentially epistemological, since it is given as a first step in his general foundationalist project. But, while it is true that Descartes' foundationalist project has been widely repudiated,<sup>2</sup> one could still deny that the *cogito* can serve as an epistemological foundation but agree that it does establish an ontological claim, namely that oneself exists as a substantial thinking subject. If so, then arguably the Buddhist non-acceptance of a permanent, substantial self which is the subject of thought would not only be false but conceptually incoherent. Of course, this 'ontological argument' for the existence of a substantial self has not by any means been universally accepted. David Hume is well-known for having raised doubt about it, which is very much in the spirit of the Buddhist perspective of non-self. In what follows, I focus on a well-known quote of the German aphorist Georg Christoph Lichtenberg (1742–1799), which captures the gist of Hume's sceptical concern about the *cogito*, and which suggests an 'impersonal' version of the *cogito*. On the other hand, Bernard Williams has argued that Lichtenberg's impersonal version of the *cogito* is conceptually incoherent (1978, 95–101). One could therefore be inclined to argue, on the basis of Williams's objection, that the Buddhist perspective of non-self is indeed incoherent. My purpose in this article is to defend the coherence of the Buddhist perspective of non-self against Williams's objection to Lichtenberg's impersonal version of the *cogito*.

In the first part of the article, I characterize more precisely the Buddhist doctrine of non-self and I underline some standard reasons why this view does seem to be coherent, at least *prima facie*, by indicating how it can accommodate some strong commonsense intuitions concerning moral responsibility and personal identity through time. The second section focuses on Descartes' *cogito* argument and how it could be used to vindicate the existence of a substantial self. I stress that the *cogito* does not require that the thinking subject must necessarily be understood as an immaterial substance. The argument could be used to support what I call 'the thesis of the substantial self', which is compatible both with Cartesian dualism and with a view, influenced by the work of P. F. Strawson, according to which the kind of self-consciousness at work in the *cogito* involves awareness of an enduring substantial self understood as a physical object among physical objects (see Strawson 1966; Evans 1982; McDowell 1994 and Cassam 1997).<sup>3</sup> In the third section, after recalling Lichtenberg's impersonal version of

2. Indeed, the most influential philosophical endeavours in epistemology in the twentieth century, in the works of such philosophers as W. Quine, W. Sellars, D. Davidson and many others, can be understood as a repudiation of Descartes' foundationalist project.

3. Of course a physical 'enduring self' would clearly end at death — it would be lasting but impermanent (hence, in Buddhist terms, be *anatta*). In this respects it would be much like the kind of self that 'Annihilationists' are portrayed as believing in, in the Pali *Nikāyas*. While these are

the *cogito*, which he uses as a way to undermine the thesis of the substantial self, I underscore two problems which it faces. I point out that the first problem can be solved by endorsing a suggestion of Derek Parfit, but that Parfit's suggestion is insufficient to solve the second problem. The second problem is tantamount to Williams's objection. Finally, I propose a general way to overcome that second problem and I conclude by pointing out that this general solution can be interpreted in two ways that differ significantly. If my solution is correct, then Williams's objection to Lichtenberg's impersonal version of the *cogito* fails to show the incoherence of the Buddhist perspective of non-self.

#### THE PRIMA FACIE COHERENCE OF THE BUDDHIST PERSPECTIVE OF NON-SELF

The Buddhist perspective of everything as non-self must be understood in the general context of the central Buddhist doctrine of the four 'noble truths', according to which our experienced world is understood as painful/stressful. Thus, the perspective of non-self, as is the case for most philosophical claims in classical Indian philosophy, is not a purely philosophical view, since it rests essentially on the soteriological motivation of achieving the cessation of that which is in any sense painful (*nibbāna*). The point is not to effect some philosophical denial of a substantial self, but encourage a rigorous letting go: an abandoning of taking anything as being, or being the possession of, a permanent, substantial 'self', based on a direct recognition that *everything* is non-self. Our existential experience of the world is understood as mentally and physically painful, in part because of a recurring experience of loss, given the impermanence of everything, and of a recurring experience of limitation, given the contingency of everything. There is a formula that is often repeated in the Buddhist *suttas* according to which a key cause of suffering is the delusion (*moha*) of not recognizing the so-called 'three characteristics' of any aspect of our experienced world, which are all impermanent, obviously or subtly painful, and non-self. Thus, according to this view, believing in the existence of a self as an enduring entity is not only a false philosophical belief but a delusional belief and a source of the recurring experience of suffering. To understand Buddhism's teaching on all as non-self, and actually to realize it in one's experiential continuum, is viewed as a rectification of that delusion, which rectification contributes to the cessation of suffering.

From a commonsense point of view, we seem to assume without much questioning that we exist as substantial enduring entities underlying our various experiences. Indeed, when we say such things as 'I was mad at him, but I no longer am', it seems plausible to think that the word 'I' is used as what some philosophers have called a device of direct reference which, in this sentence, would refer to one and the same enduring entity in both of its occurrences. Moreover, we seem to be able to make sense of counterfactual claims such as 'If I had not had such an emasculating mother, I would have had a much happier sex life'. But such counterfactual claims seemingly would make sense only on the assumption

---

sometimes talked of as simply denying any kind of self that continues beyond death (as they denied rebirth), they are sometimes also talked of as believing in a one-life substantial self, and are hence criticised by the *Nikāya* texts for believing in the 'destruction of a real being (*sato sattassa*)' — i.e. the destruction of something which, if it existed, would not end at death but be permanent, indestructible — see Harvey 1995, 28–30, 38–40.

that the word 'I' is used in both of its occurrences as a device of direct reference referring to one and the same enduring entity. Finally, a substantial conception of the self can readily explain strong intuitions we have concerning moral responsibility as well as our own identity through time. But, according to the Buddhist perspective of non-self, these commonsense intuitions are delusional, there exists no such enduring entity to which the word 'I' refers, and the failure to understand this is a source of suffering.

The Buddhist non-acceptance of a substantial self has been repeatedly developed and presented in philosophy. It is not my purpose to review these developments and the many arguments that have been developed to support it. But it is useful to recall a passage from the *Samyutta Nikāya* which indicates some central ideas in support of this perspective. The passage is particularly interesting insofar as it makes clear that this perspective is very much attuned to some of the more recent philosophical views that reject the substantial conception of the self, such as Derek Parfit's reductionist metaphysics of persons.

To properly understand this passage of the *Samyutta Nikāya*, it is first necessary to recall the Buddhist analysis of a person in terms of the five aggregates (*khandhas*). According to this analysis, which can be found in the Pali *suttas* and which occupies a prominent place in the Abhidharma literature, a sentient being is ultimately composed of five kinds of physical and psychological aggregates, namely a physical aggregate, *rūpa* or 'form', which is simply the material body, including the five senses of sight, hearing, smell, taste and touch; and four psychological aggregates: *vedanā*, that is: pleasant, unpleasant and indifferent feelings; *saññā* – perception understood as perceptual categorization; the *saṅkhāras* – 'volitional formations' in the form of desires, intentions, plans of action, dispositions to act; and *viññāna* – consciousness, the basic awareness of any object of the five senses, or of a mental object.

Within the tradition of Buddhist philosophy, much discussion and debate is concerned with the ontological status of the five aggregates, most notably in the early Abhidharma and in various philosophical schools such as the Sautrāntika, the Madhyamaka and the Yogācāra. For my purpose, it is enough to note that the theory of the five aggregates, understood simply as a conceptual analysis of what constitutes a sentient being, is quite plausible, be it only from a naïve reflective point of view, or from a phenomenological point of view.<sup>4</sup> This much should suffice to understand the following quotation from the *Samyutta Nikāya* (III.3; Bodhi 2000, 854–855):

How, householder, is one afflicted in body and afflicted in mind? Here, householder, the uninstructed wordling, who is not a seer of the noble ones and is unskilled and undisciplined in their Dhamma ... regards form as self, or self as possessing form, or form as in self, or self as in form [my emphasis]. He lives obsessed by the notions: 'I am form, form is mine'. As he lives obsessed by these notions, that form of his changes and alters. With the change and alteration of form, there arises in him sorrow, lamentation, pain, displeasure, and despair. [The same is then repeated in respect of the other four aggregates]... With the change and alteration in consciousness, there arise in him sorrow, lamentation, pain, displeasure, and despair.

4. This doctrine is typically interpreted as entailing an ontological dualism of sorts. In any case, the doctrine does seem to capture some important conceptual distinctions.

The first thing to note here is the obvious soteriological tone of the passage, which is in keeping with what I already noted concerning the essentially soteriological motivation of the perspective of non-self.<sup>5</sup> The second thing to note is that the passage clearly denies that it is acceptable to identify any of the five aggregates with 'self' in some sense; as it is repeatedly emphasized elsewhere that what is impermanent is non-self, then what is meant is a *permanent* self. It also denies that such a 'self' could be identified with something *possessing* the aggregates. Thus, it clearly rejects the claim of a 'self' that could be identified with an enduring entity that is the bearer of physical or mental properties. The latter part of the passage I italicized strongly suggests a) a denial of the claim that any of the aggregates could be viewed as a part of a whole which is 'self', and b) a denial of the claim that there is a self which is identical to a part of any of the aggregates.

It is possible to reconstruct some plausible arguments on the basis of this passage, even if such arguments are not explicitly stated in the passage. First, it seems obvious that the supposed self *qua* substantial enduring entity underlying one's physical or psychological properties cannot be identified with any of the aggregates since, given the impermanence of the aggregates, it would then not be enduring. But then, by the same token, it cannot be identified with any parts of the aggregates, nor can it be identified with a whole which would be constituted by the aggregates since such a whole would be determined by its parts and, thus, it would be no more enduring than its parts.

With respect to the claim that a putative self could be identified with something *possessing* the aggregates, things are a bit more complicated. The issue depends on how we understand the relation of possessing. If that relation is understood in such a way that *x* possesses *y* if and only if *y* are essential properties of *x*, properties which are constitutive of *x*'s identity, then again such a self would be as impermanent as the aggregates which essentially constitute it. Such a view would be tantamount to the claim of a 'self' that is a bundle of constantly changing properties, a claim often attributed to David Hume. One could, however, interpret the possession relation differently, that is, in such a way that the aggregates would only be accidental modifications of a substantial enduring entity. Such a view is committed, however, to denying that any of the impermanent aggregates are somehow constitutive of such a putative enduring self. This self becomes a mysterious entity, what some have called a 'bare particular'. While this view is not incoherent, neither is its denial.

Furthermore, there is an argument for the non-acceptance of a substantial enduring self that is very much in the spirit of Hume's well-known remark concerning the empirical unfindability of any self other than the flow of perceptions.<sup>6</sup> This argument can be used to call into question the notion of a self as a bare particular. From an empirical point of view, a principle of parsimony dictates that we

- 
5. Indeed, the passage continues by pointing out that not regarding form as self, or self as possessing form, or form as in self, or self as in form etc., there does not arise in him sorrow, lamentation, pain, displeasure, and despair (*Saṃyutta Nikāya* III.4; Bodhi 2000, 855–856).
  6. 'For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch myself at any time without a perception and never can observe anything but the perception' (Hume 1739/1951: Book I, vi, 6).

should not unnecessarily populate our ontology with entities. Thus to postulate the existence of a substantial enduring self, which has no necessary observable features — a bare particular — would seem to need a very strong argument. In other words, it seems more plausible not to postulate such an entity if we can do without it. The existence of an enduring entity underlying our physical and mental properties would seem to be vindicated only if we could not otherwise make sense of our experience. For instance, if it were impossible to make sense of our sense of personal identity through time or if it were impossible to make sense of moral responsibility without postulating the existence of a substantial self, then it might be a reason to accept it. However, as I note in what follows, the Buddhist doctrine of non-self is able to coherently account for these intuitions.

Finally, I take it that the key force of the soteriological argument, in the quoted passage, is that if one takes any of the aggregates as an enduring self, or somehow related to such a thing, one will suffer when one unwillingly becomes faced with the impermanence, and hence non-self-ness, of the aggregates. Moreover, the passage suggests that while experience is constituted only by impermanent, constantly changing aggregates, believing in the existence of a mysterious unobservable enduring entity, the existence of which is subject to doubt, might indeed be a source of sorrow and despair. Assuming that such an entity does not exist and that the impermanent aggregates do, this might be a reason to think that belief in such an entity is bound to prevent peace of mind.

Many find it intolerable not to accept the existence of an enduring self insofar as we then seem to be hard pressed to account for our strong commonsense intuitions about moral responsibility and about personal identity through time. It is not readily apparent that we can accommodate these intuitions without postulating a self as an enduring unchanging entity underlying our various experiences.<sup>7</sup> The Buddhist reply to these concerns rests on the idea that reality is an ever-changing flow of physical and mental events occurring in continuous succession, supplemented by the crucial claim that the continuous succession of ever-changing physical and mental phenomena is not random, namely it happens in a regular way.<sup>8</sup> As Rupert Gethin states:

According to the Buddhist analysis a person should be seen as [reducible to] five classes of physical and mental events that arise dependently at any given moment in time and also over a period of time. What this means then is that the causal connectedness of events is such that events occur in certain quite specific clusters and patterns ... Furthermore, causal connectedness is such that the patterns in which events occur tend to reproduce themselves and so are relatively stable over a period of time. (Gethin 1998, 142).

In other words, I view myself as the same person I was five years ago insofar as my present conscious experience, including memories, typical behavioural

7. Not to mention the issue concerning 'who is reborn' in so far as Buddhism accepts the notion of cyclic rebirth.

8. It should be noted that a similar idea has been developed at length and in many details by Derek Parfit in his reductionist metaphysics of persons. The central idea on which Parfit's reductionist view rests is that personal identity over time does not require a substantial self, it requires only psychological connectedness and/or psychological continuity. See Parfit 1984, especially p. 206. See also Siderits 2003 which underscores in many details the similarities between the Buddhist doctrine of non-self and Parfit's reductionist metaphysics of persons.

tendencies, typical beliefs and desires, and my present body occur in a *continuum* which is connected in some specific causal ways to some past conscious experience and to my body as it was constituted at that time.

Once we understand this way of accounting for personal identity over time, it becomes easy to understand how we can account for our intuitions about moral responsibility, without accepting a substantial enduring self. The perspective of non-self cannot serve as an excuse for not taking moral responsibility for our actions, or for acting without prudence in relation to the future results of our present actions. Indeed, according to the Buddhist doctrine, the present psycho-physical aggregates which actually occur are component processes in a causally-connected continuum. The present part of this psycho-physical continuum retains moral responsibility for the actions that occurred in the past in the same psycho-physical continuum, and indeed it reaps the moral consequences of these past actions. In the same way, present actions are responsible for how the psycho-physical continuum will develop in the future.

This view of personal identity as regular psycho-physical continuity is also what allows the Buddhist to say that it makes sense in our everyday transactions to use the pronoun 'I', and other similar notions, for convenience sake, as long as this does not mislead us into thinking that the word 'I' refers to an enduring entity. The use of 'I' is simply a practical expedient. This helps to understand the Buddhist idea that the doctrine of non-self does not reject a self from a conventional point of view (*samvṛti*) but only from an ultimate point of view (*paramārtha*). Indeed the word 'self (*atta/ātman*)' is seen as acceptable in the *Nikāyas* when used simply to mean 'oneself', that is the empirical self constituted by the impermanent aggregates which are non-self, in the sense of not a permanent, substantial self.

These remarks suggests that we can indeed make coherent sense of our experience without postulating the existence of 'self' as a bare particular, and parsimony dictates that we should deny that there exists such a 'possessor' of form, feeling, etc. This establishes the *prima facie* case for the coherence of the Buddhist doctrine. In the next sections, though, I discuss an argument that stems from Descartes' *cogito* argument which, if correct, would entail that the non-acceptance of a substantial enduring self is conceptually incoherent.

#### THE BUDDHIST NON-ACCEPTANCE OF A SUBSTANTIAL ENDURING SELF MEETS DESCARTES' *COGITO*

Let us start by briefly revisiting Descartes' notorious *cogito* argument. If the *cogito* argument establishes anything, it establishes that a self as a substantial agent of thought does exist, and this is in direct contradiction with the Buddhist perspective of non-self. As I noted, Descartes' motivation for the *cogito* was essentially epistemological, since this argument is given as a first step in his foundationalist project. Descartes seeks something he would know infallibly, to serve as the ultimate foundation of knowledge. To arrive at such an infallible foundation, Descartes imagines a scenario where he doubts just about everything he previously believed, because he reasons that if in such a context there is one thing which he cannot doubt, then he will have found the infallible knowledge he seeks. This is what is known as the method of hyperbolic doubt. Descartes deploys this method by imagining that, contrary to what we normally think, all his percep-

tions would not be produced by objects in the external environment, but by a powerful evil demon deceiving him. Descartes arrives at the conclusion that even if such a conceptually possible scenario would be the case, then there is still one thing which he could not doubt, namely that he thinks and exists. In this contexts, Descartes concludes: ‘the proposition “I am, I exist” is necessarily true every time I utter it or conceive it in my mind’ (1641/1993, 18), and thus he claims to have established an ontological point on the basis of purely conceptual thinking: that at least one thing exists namely himself qua thinking subject. This point is quite clear in the following passage from the ‘Second meditation’:

But what then am I? A thing that thinks. What is that? A thing that doubts, understands, affirms, denies, wills, refuses, and that also imagines and senses. ... Is it not *the very same ‘I’* who now doubts almost everything, who nevertheless understands something, who affirms that this one thing is true, who denies other things, who desires to know more, who wishes not to be deceived, who imagines many things even against my will, who also notices many things that appear to come from the senses? What is there in all of this that is not every bit as true as the fact that I exist. (Descartes 1641/1993, 20.) [My emphasis]

This ‘thing that thinks’, the existence of which Descartes claims to know infallibly, must be understood as an entity in a substantial sense, that is: ‘this very same “I”’ must be understood as something which endures in the continued succession of one’s experience, an enduring existent to which the word ‘I’ refers when one says, or thinks ‘I exist’. This suggests that the word ‘I’ is indeed a device of direct reference.

It is important to note that at this stage of the argument, while Descartes claims to know infallibly that he exists in a substantial sense, he is not saying anything yet about the immaterial nature of this thing. This point is important, because it indicates that the *cogito* can be vindicated by philosophers who hold that the word ‘I’ is indeed a device of direct reference and that it refers to a physical object among other physical objects.<sup>9</sup> Indeed, as is well known, in order to hold that the subject is an immaterial substance, Descartes uses a further argument. According to this argument, which we may call ‘the argument for dualism’, a thinking subject could not be a material substance because, while it is possible to doubt the existence of a material substance, it is impossible to doubt the existence of a thinking subject. Many have noted, however, that this argument is fallacious insofar as it claims to establish a metaphysical conclusion concerning the nature of the thinking subject on the basis of some purely epistemological premises. The upshot is that while one could accept that the certainty of the *cogito* establishes the existence of the thinking subject, this does not entail anything whatsoever about the nature of that thing. Thus there is a sense in which one could hold that the *cogito* establishes the existence of the thinking subject in the substantial sense that the word ‘I’ does refer to an enduring object, without being thereby committed to substance dualism. I call this claim ‘the thesis of the substantial self’. There may be various ways of interpreting this thesis, but on any of its interpretations it obviously contradicts the Buddhist non-acceptance of an enduring, substantial self, whether of a kind that lasted for one life or for many.

9. Here I think particularly of the view of Quassim Cassam, who develops some suggestions of Peter Strawson (see Strawson 1966; Evans 1982; McDowell 1994 and Cassam 1997).



Thus, if the *cogito* supports the thesis of a substantial self, not only is the Buddhist view on this false, but it is conceptually incoherent. To see this we must recall that the *cogito*, and hence the thesis of the substantial self, is supposed to be established by the application of the method of hyperbolic doubt, that is, the evil demon scenario. While the general criteria of conceptual coherence may remain an open question, it is nevertheless quite plausible to hold that if something is indubitably true then its negation is conceptually incoherent, in other words we cannot coherently conceive its being true. Therefore, if the *cogito* argument, which is stated in the context of the hyperbolic doubt, supports the thesis of the substantial self, it supports it as something which is indubitable and, thus, the Buddhist non-acceptance of such a self would not only be false but conceptually incoherent. This consequence is quite strong: it would establish that what is a delusion is not the belief in the existence of a substantial self, as the Buddhist would have it, but to believe that there are no grounds for accepting such a thing.

As we know, the thesis of a substantial self has not been universally accepted. In addition to Hume's objection to it in the *Treatise on Human Nature*, Kant also rejects it in the *Critique of Pure Reason*, claiming that the *cogito* does not entail the existence of a thinking subject in a substantial sense, but that it entails only the necessity of a transcendental ego, whatever that may mean. Moreover, contemporary philosophers developing ideas of Wittgenstein have denied that the I-concept is a referential concept. However, in what follows, I want to focus on an objection to the thesis of a substantial self which is very much in the spirit of the Buddhist perspective of non-self. This objection was raised by the German aphorist Georg Christoph Lichtenberg. In a famous passage, Lichtenberg formulates what we may call an impersonal version of the *cogito*. As he puts it:

We are acquainted only with the existence of our sensations, imaginations, and thoughts. 'Thinking is going on' (*Es denkt*) is what one should say, just as one says, 'Lightning is occurring' (*Es blitzt*). Saying '*Cogito*' is too much, as soon as one translates it as 'I am thinking'. Accepting, postulating, the *I* is a practical requirement. (1971, 412, §76)

As I understand it, Lichtenberg's remark indicates that saying 'I think' would be saying too much in the sense that it would be saying more than what the context of the hyperbolic doubt warrants. It also suggests that in conscious experience and conscious thinking we are conscious only of the process and content of experience and thought; we are not conscious of a subject, and such conscious experiences and thoughts do not entail the existence of a thinking subject. Talk of 'I' is just a pragmatic necessity of discourse, with no ontological implications. This view, which we may call 'the thesis of selfless thinking', obviously contradicts the thesis of the substantial self, and it is quite similar to the Buddhist perspective of non-self.

Lichtenberg's impersonal account of the *cogito* faces two serious problems. The first is that it is highly counterintuitive to hold that in the context of a conscious reflection such as the one Descartes suggests, the use of the first person pronoun 'I' would be dispensable. On the contrary, it seems to be crucial to the understanding of Descartes' meditation that it must be done from the first person perspective, that is: precisely by using the I-concept. And it seems counterintui-

tive to hold that we could say ‘Thought is going on’, instead of ‘I think’.

To solve this problem, we can use a suggestion of Bernard Williams which has been vindicated by Derek Parfit (1984, 225). According to this suggestion, the first person pronoun ‘I’ must indeed appear when we re-enact Descartes’ reflections in his ‘Second meditation’. But, we can say that the pronoun ‘I’ is only mentioned to identify the content of the conscious thinking which occurs, and that it is not used to refer to an enduring entity. Thus, the thesis of selfless thinking can overcome this first problem by describing what is going on in Descartes’ *cogito* argument, not by saying ‘I think’ but by saying the following instead: ‘It is thought: “I think”’. This suggestion solves the first problem, because it acknowledges that what is going on is indeed the occurrence of what we may call a subjective thought, as Parfit notes, ‘since the subject of experiences is here mentioned only in the content of the thought, this sentence does not ascribe this thought to a thinker’ (1984, 225). This suggestion is very much in the spirit of Lichtenberg’s remark, according to which we are acquainted only with the process and content of experience and thought. The upshot is that this first problem is not very serious. The second problem, however, is much more serious.

The second problem is that the statement ‘There is thinking going on’ is very general and somewhat vague. In a reflection such as the one suggested by Descartes, it seems that we are certain of something a bit more specific. As Bernard Williams’s discussion of Lichtenberg’s impersonal version of the *cogito* suggests, we should be able to distinguish between cases where two contradictory thoughts, for instance ‘I doubt that *p*’ and ‘I do not doubt that *p*’ simply express a disagreement between two persons, or the fact that a person has changed his or her mind at a later time, and cases where they express that a person is being logically incoherent (Williams 1978, 95–101). Lichtenberg’s impersonal version of the *cogito*, however, is not able to account for such a distinction. The problem is that the impersonal ‘It is thought: “I think”’ lacks an essential indexical anchoring. And, on this basis, there is only a short step to hold that only a referential use of the word ‘I’, referring to a substantial self, can provide the required indexical anchoring. This would entail that the thesis of selfless thinking is conceptually incoherent.<sup>10</sup>

In light of this problem, Quassim Cassam has pointed out that the ascription of first person thought as in ‘It is thought: “I think”’ requires a form of relativization (1997, 185). Cassam recalls Parfit’s suggestion, which he calls a ‘weak relativization’, according to which the thought could be relativized by indexing it on the body on which it depends or on the particular life continuum which contains the occurrence of the thought expressed by ‘I think’ (1984, 225–226). Cassam uses Williams’s problem as a way to reject the coherence of the thesis of selfless thinking, in order to vindicate what he calls a ‘strong relativization’, in support of his own materialist version of the thesis of the substantial self (1997, 183–196). The main problem with Parfit’s and Cassam’s suggestions is that they rest on metaphysical assumptions that go beyond the restricted epistemological context of Descartes’ hyperbolic doubt. Parfit assumes the existence of a physical body or

10. Here one might object that William’s objection requires the assumption that other persons – other thinkers – exist, which already goes beyond Descartes’ initial position of radical doubt. However, he only needs to posit that without some indexical anchoring, it is unclear if the relevant thought is had by Descartes himself or some agent or agents (whether an ‘evil demon’ or e.g. a brain-manipulating scientist) that might be deceiving him.

of a psycho-physical continuum extending beyond the context of the occurrence of 'I think', and Cassam must obviously assume that physical bodies exist. But the context of Descartes' hyperbolic doubt dictates precisely that we should suspend our judgment about the existence of such things. I think, however, that there is a much simpler way to address the problem.

#### A SIMPLE SOLUTION TO WILLIAMS'S PROBLEM

As I noted, Williams's problem arises because the impersonal version of the *cogito* 'It is thought: "I think"' would seem to lack an essential indexical anchoring. To see this more precisely, suppose that at a certain time *t*, the two following propositions are true:

- (1) It is thought: 'I doubt that *p*'
- (2) It is thought: 'I do not doubt that *p*'.

It would seem that

- (3) It is thought: 'I doubt that *p* and I do not doubt that *p*'

is not entailed by the conjunction of (1) and (2) in so far as it is left open that (1) and (2) occur separately. Thus, (1) and (2) could simply express a disagreement between two people about whether or not *p* is doubtful. If, however, (1) and (2) definitely do not entail (3), it must be assumed that (1) and (2) do indeed occur separately. But nothing in the impersonal formulation determines whether they occur separately or not. In other words, without some kind of indexing of (1) and (2) that would make it clear whether or not they occur separately, it is left open that a self-contradictory thought might occur. Without such an indexical anchoring, the possibility of (3) would be left open.

Thus, Lichtenberg's impersonal version of the *cogito* might seem to be incoherent, insofar as the only way in which first person thoughts can be indexically anchored is by a referential use of 'I' referring to an enduring self. While Lichtenberg's impersonal version of the *cogito* leaves it indeterminate whether the occurrences of (1) and (2) express only a disagreement between two persons or the fact that a person has changed his or her mind at a later time, or whether they express a contradiction, it is clear that they do express a contradiction if 'I' is used referentially to refer to one and the same subject at one and the same time.

If this is correct, then Lichtenberg's impersonal version of the *cogito* would indeed be incoherent, and the *cogito* argument would entail that the thesis of selfless thinking is conceptually incoherent, and hence that the Buddhist perspective of non-self is conceptually incoherent. But, in order to hold, on this basis, that the thesis of selfless thinking is incoherent, one must argue that *only* a referential use of 'I' can provide the required anchoring. This, however, is far from obvious. Indeed I can think of two other alternatives, neither of which entails the thesis of a substantial self.

The general strategy underlying these two alternative ways of providing the required indexical anchoring consists in claiming that in the impersonal formulation of the *cogito*, where 'I' is only mentioned, there is a hidden indexical. Making this hidden indexical explicit, we could schematically represent the impersonal formulation in the following way: 'It is thought<sub>[ij]</sub>: "I think"'. The idea is simply that it is coherent to think that the very occurrence of the thought itself can

provide the indexical anchoring. In other words, it is plausible that in a *cogito*-like context of thought, there is an indexical reference to the very occurrence of the thought.

The first interpretation of this general strategy is inspired by so-called Higher-order theories of consciousness, various versions of which have been suggested in the literature.<sup>11</sup> For instance, according to David Rosenthal's influential Higher-order thought theory of consciousness, the occurrence of a conscious thought, say *O*, at time *t*, (say the thought that I doubt that *p*) is conscious only if it is accompanied, at time *t* or just before *t*,<sup>12</sup> by the occurrence of an unconscious second-order thought, say *O'*, the content of which is that the first-order thought *O* is occurring. Rosenthal's theory can thus be understood as claiming that the second-order thought *O'* is precisely an impersonal thought which contains a demonstrative indexical referring to the occurrence of the first-order thought *O*. In other words, on this interpretation, the impersonal version of the *cogito* could be stated like this: 'It is thought <sub>[this]</sub>: "I doubt that *p*"', where the hidden demonstrative '[this]' provides the indexical anchoring that was missing. On such a view,

- (1') It is thought<sub>[this]</sub>: 'I doubt that *p*', and  
 (2') It is thought<sub>[this]</sub>: 'I do not doubt that *p*',

do not entail

- (3') It is thought<sub>[this]</sub>: 'I doubt that *p* and I do not doubt that *p*'.

This is because the propositions expressed by (1') and (2') are made determinate simply by the context where they occur. It is easy to see that these two sentences can very well occur in two different contexts, where the two occurrences of '[this]' refer to two different thoughts. For example, (1') could occur in a particular mind stream while (2') occurs in a separate mind stream. Or again, (1') could occur in a particular mind stream at a certain time, while (2') might occur in the same mind stream but at a different time. Thus, it is possible that (1') and (2') are true, while (3') is false.

Is it left open that (1') and (2') occur in the same context which would mean that the two occurrences of '[this]' refer to one and the same thought? First, it should be noted that, generally speaking, it is not ruled out that two occurrences of a demonstrative may occur in the same context, referring to one and the same thing. Suppose that my friend and I are looking at a painting of Picasso in a museum and I say: 'This is beautiful'. My friend may agree and say: 'This is

11. See for instance, Rosenthal 1997, Lycan 1996, Gennaro 2004, and Carruthers 2005. Higher-order theories of consciousness are somewhat in the spirit of the view held by the Nyāya philosophers, in classical Indian philosophy — see Matilal 1986, 141–179.

12. The reason why the unconscious accompanying thought (or quasi-perception) cannot occur *after* the conscious mental event, is that the purpose of these theories is to *reductively* explain consciousness, that is to explain the nature of conscious mental events in terms of something else. That is, they try to answer the question: By virtue of what is a conscious mental event conscious? The answer is that it is conscious in so far as it is accompanied by an unconscious mental event which is intentionally directed at the conscious mental event. Such an approach is not without its problems, but it is clear that if it is to be a plausible philosophical account of the nature of conscious mental events, then the accompanying unconscious mental event *which makes the conscious event conscious* cannot occur after the conscious event, for this would leave it open that at time *t* there is a conscious mental event while what is supposed to make it conscious does not exist (yet).

beautiful indeed'. But matters seem to be quite different in the scenario that concerns us here. Prima facie, such a scenario seems to be ruled out insofar as the occurrence of '[this]' in (1') refers to 'I doubt that *p*', while its occurrence in (2') refers to 'I do not doubt that *p*', which are obviously two different thoughts. Insofar as two different thoughts are referred to, then the two occurrences of '[this]' must occur in two different contexts. A scenario where (1') and (2') occur in the same context and the two occurrences of '[this]' refer to one and the same thought seem extremely difficult to conceive. The only way to conceive it, and it is unclear whether this is even conceivable, is if we already assume that it is possible that the thought (or quasi-thought, if I may say) expressed by (3') can occur in a mind stream, at a certain time. If so, then the occurrences of '[this]' in (1') and (2') could be said to occur in the same context, namely in the context where the thought (or quasi-thought) expressed by (3') occurs. In such a case the two occurrences of '[this]' in (1') and (2') could be said to refer to two different parts of one and the same thought (or quasi-thought), namely the thought (or quasi-thought) expressed by (3'). As I noted, it is unclear whether such a scenario is conceivable. In any case, if it was conceivable then Williams's objection would simply not arise, insofar as this objection rests precisely on the assumption that contradictory thoughts, like the one expressed by (3') are impossible.

The second way in which this general suggestion can be interpreted is by claiming that conscious thoughts are *reflexive* in the sense that they are self-referential.<sup>13</sup> According to this reflexive interpretation, the hidden indexical in 'It is thought<sub>[ij]</sub>: "I think"' is not understood as a demonstrative, but only as making explicit the reflexive, or self-referential, character of the conscious thought. On this interpretation, the impersonal version of the *cogito* could be understood as follows: 'It is thought in *this very thought*: "I think"'. It is easy to understand that this suffices to resolve the problem of the indeterminacy in (1) and (2).

There may be another way to solve Williams's problem.<sup>14</sup> Thoughts seem to be something that can only really occur in a network, in relation to other thoughts, in the sense that any specific thought is lent meaning and significance by other thoughts that a person may have. This idea, which corresponds to the doctrine of semantic holism, suggests that the network of thoughts of which specific thoughts are a part could perhaps provide the required anchoring. But how could that be? How is it that since (1) and (2), occurring at time *t*, have meaning

13. This idea is congenial with so-called self-representational theories of consciousness, which have recently been proposed in contemporary philosophy of mind — see for instance Kriegel and Williford 2006 and Kriegel 2009. In classical Indian philosophy, a reflexive account of consciousness (*svasaṃvedana*) is notoriously held within the tradition of Dignāga and Dharmakīrti. Whether conscious mental events are literally reflexive is an important issue debated in the Abhidharma literature. Yao 2005 traces the historical genesis of Dignāga's and Dharmakīrti's reflexive view to the Mahāsaṅghika school and also possibly, in part, to some developments within the Sautrāntika tradition. He also makes it clear that Buddhaghosa's commentary on the *Kathāvatthu* (XVI.4) strongly suggests that the Theravāda school rejects reflexive awareness, if it is strictly understood as one *citta* (i.e. one consciousness) referring to itself. In that passage, Buddhaghosa accepts, however, that *one present moment* of consciousness can be said to be reflexive in a loose sense, insofar as a 'present moment' can be subdivided and, thus, a *citta* is known not by itself but by another numerically distinct *citta* — see Yao 2005, 25–30. Chapter 3 of that book also makes it clear that the Sarvāstivāda school argues against reflexive awareness.

14. This was suggested by an anonymous referee.

only by virtue of being embedded in networks of thoughts, the possibility of (3) is ruled out?

The answer that comes to mind is that whatever might be the network of thoughts in which (1) is embedded, then at time  $t$ , it should not contain a thought which is a logical contradiction of (1). Thus, it may seem that (2) could not be part of that network and, thus, while (1) and (2) could be true, at  $t$ , (3) would be ruled out. But whether (2) contradicts (1) depends on the logical form of 'I doubt that  $p$ ' and of 'I do not doubt that  $p$ '. If 'I' was a logically proper name — a name which cannot but refer to one and the same thing — then these two sentences would be the negation one of the other, and it would be ruled out that both are part of the same embedding network of thoughts. But 'I' is not a logical proper name and it remains unclear that (2) does contradict (1). Hence, it remains unclear why the network which determines the meaning of (1), at  $t$ , could not contain (2). After all, it is plausible that the network which determines the meaning of 'John is tall' might contain 'Bob is not tall', 'That guy is not tall', or 'Someone is not tall', etc. It remains unclear how the network could provide the required anchoring.

What is the logical form of 'I doubt that  $p$ ' in (1)? One might say that 'I' is an indexical device of direct reference, which can refer to different enduring selves. If this is the only possible answer, then the existence of enduring selves follows. One could however hold that the occurrences of 'I' in 'I doubt that  $p$ ' and 'I do not doubt that  $p$ ' are different disguised definite descriptions. Thus, (1) and (2) could be analysed as

- (1\*) It is thought: 'The one and only mental continuum M doubts that  $p$ '
- (2\*) It is thought: 'The one and only mental continuum Q does not doubt that  $p$ '.

Accordingly, while the meaning of (1\*), at  $t$ , is determined by a network of thoughts, it is ruled out that such a network contains:

- (3\*) It is thought: 'The one and only mental continuum M doubts and does not doubt that  $p$ '.

Thus, the network of thoughts would provide the anchoring required in a way which is consistent with the Buddhist non-acceptance of an enduring self.

This solution, however, is a bit schematic. To make it more precise we would have to give a more specific description of the one and only mental continuum M, making clear that  $M \neq Q$ . Though this is not something I want to argue at length here, it seems that the definite description which succeeds in picking out the specific mental continuum would need to make some kind of indexical reference to the occurrence of the specific thought, for instance (1) at time  $t$ , which is part of a larger network determining its meaning. As far as I can tell, such an indexical reference could only take one of the two forms I have suggested, namely the reflexive approach or the demonstrative interpretation of the Higher-order theory.

To sum up, Descartes' *cogito* argument could be used to support the incoherence of the Buddhist non-acceptance of an enduring substantial self only provided that Williams's argument shows that Lichtenberg's thesis of selfless thinking is conceptually incoherent. But, if my argument is correct, the latter is perfectly coherent, and hence the Buddhist perspective of non-self remains intact.

## CONCLUSIONS: REFLEXIVITY OR HIGHER-ORDER?

I would like to conclude by clarifying further the difference between the two interpretations of the general solution I propose. It should first be stressed that these two interpretations are significantly different. According to the reflexive or self-referential interpretation, the impersonal version of the *cogito* 'It is thought: "I think"' is understood as expressing the occurrence of only one mental event, namely a conscious mental event. But according to the interpretation of the Higher-order theory I suggest, it expresses the occurrence of two different mental events which are linked by the hidden demonstrative occurring in the first of these two mental events, which is itself unconscious.

In so far as we are talking about an impersonal version of the *cogito*, it should be stressed that 'It is thought: "I think"' is taken to express what is occurring in the context of hyperbolic doubt. If this is something that is known infallibly, then it would mean that the reflexive interpretation is the correct one. This is because in the reflexive interpretation, the unique conscious thought is at once what does the conceptual indexing, or the referring, and what does the ontological anchoring, or what is referred to, which rules out failure of reference. The occurrence of the thought guarantees that it refers to something, namely itself. Moreover, this is very much in the spirit of Lichtenberg's remarks. When he notes that Descartes is saying too much when he implies the existence of a self, we could add that he is indeed saying too much in the sense that, in the context of hyperbolic doubt, all he can be certain of is the occurrence of the thought.

On the other hand, in the first interpretation, the impersonal version of the *cogito* 'It is thought: "I think"' is understood as expressing the occurrence of two distinct mental events: a second-order unconscious mental event that contains the hidden demonstrative, namely 'It is thought <sup>[this]</sup>' and a first-order conscious mental event that is demonstratively referred to by the hidden demonstrative, namely 'I think'. Given that the relation between these two mental events is contingent, it is not necessarily ruled out that such a relation might fail to obtain. In other words, it is not necessarily ruled out that the second-order thought might occur while the hidden demonstrative is vacuous, that is, that it fails to refer to anything. Thus, the second interpretation leaves open the possibility of a very strange case. In such a case, the result would be that there is only the illusion of the occurrence of a conscious mental event! We may wonder whether such a possibility is intelligible. Thus, on this interpretation, the result of the hyperbolic doubt would not be that it is certain that a conscious mental event occurs but that nothing whatsoever is certain.

Moreover, to the extent that the impersonal version of the *cogito* is to be understood strictly within the epistemological context of the hyperbolic doubt, it would seem illegitimate to postulate the existence of an unconscious second-order mental event demonstratively referring to the first-order conscious mental event. This is because the hyperbolic doubt dictates that we suspend judgement concerning what is not present to consciousness. These remarks indicate that the reflexive interpretation should probably be preferred.<sup>15</sup> Be that as it may, if

15. It should be noted, especially for readers acquainted with the practice of active mindfulness (*sati*), that normally the self-reflexive character of a conscious mental state is quite naturally less noticeable than when one practices active mindfulness.

what I have argued is correct, the main conclusion to draw is that the Buddhist non-acceptance of an enduring substantial self is indeed coherent despite the problem raised by Williams against Lichtenberg's impersonal version of the *cogito*. Moreover, parsimony strongly suggests that it should be preferred to the claim that there exists an enduring subject underlying our various experiences.

#### ACKNOWLEDGEMENT

I would like to thank Jay Garfield, Jeremy Hayhoe and an anonymous referee for their comments on previous versions of this article. I would also like to thank Michael Drummond.

#### BIBLIOGRAPHY

- Bodhi, Bhikkhu. 2000. *The Connected Discourses of the Buddha. A Translation of the Saṃyutta Nikāya*. Boston, MA: Wisdom Publications.
- Bodhi, Bhikkhu, ed. 1993. *Abhidhammattha Sangaha. A Comprehensive Manual of Abhidhamma*. Onalsaka, WA: BPS Pariyatti Editions.
- Carruthers, P. 2005. *Consciousness: Essays from a Higher-Order Perspective*. Oxford: Oxford University Press.
- Cassam, Q. 1997. *Self and World*. Oxford: Clarendon Press.
- Descartes, R. 1641/1993. *Meditations on First Philosophy*, translated by D. A. Cress. Indianapolis, IN: Hackett.
- Evans, G. 1982. *The Varieties of Reference*, edited by J. McDowell. Oxford: Oxford University Press.
- Gennaro, R., ed. 2004. *Higher-Order Theories of Consciousness*. Philadelphia, PA: John Benjamins.
- Gethin, R. 1998. *The Foundations of Buddhism*. Oxford: Oxford University Press.
- Harvey, P. 1995. *The Selfless Mind — Personality, Consciousness and Nirvana in Early Buddhism*. London: Routledge.
- Hume, D. 1739/1951. *A Treatise of Human Nature*, edited by I. A. Selby-Bigge. Oxford: Oxford University Press.
- Kant, I. 1929. *Critique of Pure Reason*, translated by N. Kemp Smith. London: Macmillan.
- Kriegel, U. 2009. *Subjective Consciousness*. Oxford: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199570355.001.0001>
- Kriegel, U., and K. Williford, eds. 2006. *Self-Representational Approaches to Consciousness*. Cambridge, MA: MIT Press.
- Lichtenberg, G. C. 1971. *Schriften und Briefe*. Vol. 2. Munich: Carl Hanser.
- Lycan, W. 1996. *Consciousness and Experience*. Cambridge, MA: MIT Press.
- Matilal, B. K. 1986. *Perception*. Oxford: Clarendon Press.
- McDowell, J. 1994. *Mind and World*. Cambridge, MA: Harvard University Press.
- Parfit, D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Rosenthal, D. 1997. 'A theory of consciousness', in *The Nature of Consciousness*, edited by N. Block, O. Flanagan, and G. Güzeldere, 729–754. Cambridge, MA: MIT Press.
- Siderits, M. 2003. *Personal Identity and Buddhist Philosophy*. Aldershot: Ashgate.
- Strawson, P. F. 1966. *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason*. London: Methuen.
- Williams, B. 1978. *Descartes. The Project of Pure Inquiry*. London: Penguin Books.
- Yao, Z. 2005. *The Buddhist Theory of Self-Cognition*. New York: Routledge.