PHILOSOPHICAL PERSPECTIVES

Philosophical Perspectives, 00, 2018
doi: 10.1111/phpe.12098

SELFLESS EXPERIENCE

Geoffrey Lee
University of California, Berkeley

Hume famously said that you are "*nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement*"[1]. One well-known objection to this "bundle theory of the self" is that there is no way to understand what a perception or an experience is without appealing to the notion of a subject or self. An experience is something that essentially involves a self having the experience, and therefore, according to the objection, selves are prior to experiences[2]. Compare the claim that the existence of electrons is no more than the existence of certain bundles of "electric charge events". If the relevant electric charge events exist in virtue of electrons existing (and being charged), this "bundle theory of electrons" can't be correct.

My primary aim in this paper is not to defend a bundle theory of the self, but to look at the claim that we can't understand what conscious experiences are without appealing to a self. I will explore a conception of experiences on which they are *not* most fundamentally understood as enjoyed by selves or subjects, and on which the existence of experience is either prior to, or independent from, the existence of selves[3]. In so doing I am following in particular Parfit (1999) who has argued forcefully for a set of views quite similar to those I will defend in this paper (although he doesn't explicitly consider the main thesis I'll be exploring here).

The view I will defend could be regarded as a kind of "No Self" view, but we should be very careful about what we mean by denying the existence of "the self". Eric Olson (1998) complains that there is no "problem of the self", on the grounds that different authors purporting to discuss the existence or nature of "the self" often have in mind quite different kinds of things, and therefore risk talking past each other in debating about "the self". For example, a theorist who thinks of the self as the subject of conscious awareness might be talking about a quite different entity from a theorist who is primarily interested in selves as moral or political subjects. I'm sympathetic with Olson's critique of the existing literature, although I think that his conclusion is perhaps a little overstated. The right conclusion is that there are many closely connected conceptions of what a "self" is, and that for each such conception there is a separate problem

of figuring out what, if anything, satisfies this conception. So there are many connected problems of the self, not *no* problem.

Furthermore, I think that when someone uses "self" without making explicit a particular self-conception, we shouldn't assume that there is some pre-theoretical notion of "self" that they clearly have in mind; rather we should treat what they are saying as poorly defined. I take a self-conception to be a description that a thing must satisfy to be a self, or a "self-role"—a specification of the role that a self is supposed to play in a theory of, say, the conscious mind, or some other subject matter. Some authors (e.g. Strawson (1999), Bayne (2010)) do specify self-roles in discussing the self; but the tendency is to specify multiple roles at once, and not consider independently what, if anything, satisfies each role. This will not be my approach here, indeed I will be focusing on one role in particular—the self as metaphysical subject of experience: the thing that has experiential properties[4]. I see this role as central to many debates about the self, especially those start by thinking of the self "subjectively", as at the center of a conscious life. (We might contrast this with an "objective" conception, on which the self is a thought of as a physical object like a human organism. On that conception, facts about conscious experience may be less central in individuating selves. I do not see these conceptions as competing in a substantive way; at most they involve interest in different self-roles). The thesis I will discuss is that there is no self, understood as metaphysical subject of different experiences in a conscious life.

Admittedly, other self-roles, such as the self as the object of self-awareness, or the object of self-reference (the "I"), are also important in debates about the self, including debates where we are thinking of the self "subjectively". As I will argue, these roles are not a priori equivalent to the metaphysical subject role, and may even be played by different objects. Moreover, these other self-roles probably *are* satisfied. Therefore, since some theorists may put more weight on them than on the metaphysical subject role, they may not take my thesis to show that there is no "self". The thesis is nonetheless significant for such theorists in a different way, because it suggests that experiences are individuated independently of the "self", as they understand it. In so far as the existence of this "self" even partly consists in the existence of a conscious life, this will give us something like a Humean bundle theory: experiences exist independently of selves, and selves are (at least partly) constituted by experiences. Thus, assessing the main thesis of this paper is of central importance to the traditional Humean debate about the existence and nature of the self (see 5.1 and 6 below for more discussion).

Bearing this in mind, I will begin by explaining in detail how I am conceiving of the metaphysical subject role (section 1), arguing that it is a substantive question what the metaphysical subject of experience is (section 2). I will then give a precise formulation the main thesis I'm interested in here, which I call the "Subject Non-Identity Thesis" (section 3). I'll give some positive motivations for the thesis (section 4), before responding to some objections (section 5).

1. Metaphysical Subjects and the Subject Non-Identity Thesis

As I said, I am interested here in the "self" conceived of as the metaphysical subject of a mental event. What is a metaphysical subject? It is a component of an event. I will assume a view of events on which they involve something happening to an object or group of objects. My dancing involves *me* dancing, a dance performance involves a group of dancers dancing (and maybe other objects like a theater and audience). In general, I will assume that an event is at least partly individuated by an object or group of objects, and a way in which they are propertied and related (perhaps relative to a time or over a period of time). The metaphysical subject (or subjects) of an event are simply those objects that individuate the event by instantiating the relevant properties and relations. (An example of a view on which events clearly have metaphysical subjects is Kim's well known (1976) view of events as property instantiations; I suspect the notion will have application on almost any reasonable view of events, however).

Mental events—in particular, experiences, which I will be focused on—also have metaphysical subjects. For example, if I feel a sharp twinge in my elbow, then the metaphysical subject of this event might be me, and the relevant property the property of feeling a sharp elbow pain. So the metaphysical subject of an experience is simply the item that individuates the experience by instantiating an experiential property (assuming the experience has a single metaphysical subject (see below)). The metaphysical subject of experience, in this sense, is clearly a central strand in our notion of the self (although not the only strand (see section 5.1)). As I will argue though (section 2), we shouldn't just assume that the metaphysical subject of an experience is a person like me: it might be a part of me, like my brain, or a part of my brain, or a more esoteric object like a temporal part of a brain-part.

I'm assuming here that an experience has a single metaphysical subject. One could hold instead that experiences are really plural events: events involving something happening to a group of objects. For example, someone might hold that an experience consists in a group of neurons each firing in a particular way, and all influencing each other in a particular way, much as a dance routine might involve a group of dancers each dancing in a particular way, each influenced by the movements of the other dancers. Thus a pain might really be the event of a group of neurons being "arranged pain-wise"; more generally, instead of having a single metaphysical subject instantiating an experiential property, maybe we instead have a group of subjects "arranged experience-wise". Indeed, on the view I will argue for, total experiences are helpfully thought of as plural events (although not quite in the way just described).

Having said this, it is plausible that plural events are equivalent to singular events, in a certain sense. We can treat a plural event as equivalent to a singular event involving an object that is a mereological fusion of the relevant plurality, enjoying a certain structural property. For example, a dance routine can be understood as involving a single object—the dance troupe—having certain dancers

at parts, and these parts being propertied and related in a particular way. Or an experience might involve a single object, a brain area, being a fusion of a group neurons, and the neurons being propertied and related in a certain way[5]. Thus we can unproblematically talk about experiences as if they have a single metaphysical subject. In what follows I'll talk this way, although nothing turns on this.

Some events involve an intuitive but vague distinction between objects that are the central focus of the event—the items to which something is happening—and other objects whose existence may be required for the event to exist, but are not the main focus of the event. For example, in a game of football, the players, the ball, and the place where the game happens are the main focus of the event. But for it to be a game of football, the whole institution of football has to exist, which involves many objects spread out over history. Fortunately, this distinction is not as important in the case of experiences, because they do not have highly extrinsic enabling conditions in the same way, or so I will assume. In particular, I'll assume that experiential properties are either intrinsic properties, or if they are relational, that they are relations to the *targets* of experience, such as medium sized objects like chairs (see 4.3 for more discussion). On the intrinsic view, the metaphysical subject of experience (including its parts) is the only object individuating the experience, so we won't have to draw an arbitrary distinction between focus and periphery. Things are a little more complicated on the relational view: on these views, there is a natural distinction between the item instantiating the experiential property (the metaphysical subject), and the target items to which it is experientially related; and there may be other objects that enable the relation to hold, such as photons bouncing off the object. So the metaphysical subject will *not* be the only object involved in individuating the experience. Also note that if we combine a relational view and a plural view, an experience will involve a group of objects (such as neurons) being jointly related to the target of the experience. (See section 5.5 for more discussion).

Factors other than a metaphysical subject (or subjects) and experiential property may be involved in individuating an experience. For one, the property instantiation may be relativized to a time or period of time (this may not be required if the metaphysical subject is a short-lived temporal part of a 4D object). Second, we seem to individuate events (including mental events) in terms of the determinate way in which the relevant property instantiation is realized. For example, if at time t I'm dancing by walzing, but I could have instead been dancing at t by lindy-hopping, then intuitively this would have been a different dancing, even though it involves the same property, subject and time. This is an important way in which events may differ from facts: in both cases the fact that I am dancing at t is the same, even though we have a different dancing in each case. (It is also a way in which a reasonable view of event individuation might differ from Kim's (1976) view, mentioned above). None of this matters for my purposes, however: all I need is that there is an object (or objects) and experiential property that *partly* individuate an experience.

As well as talking about individual experiences, I will take for granted here the notion of a "conscious life" or "stream of consciousness": a process which has as parts the conscious experiences that occur in the life of an ordinary person. I will assume that a conscious life is unified in the sense that these experiences don't involve such exotica as branching consciousness, or other radical forms of disunity; I do not require that a conscious life is "unified" in any stronger sense.

Although I will be arguing for what could be regarded as a kind of "no self" view, I won't be denying that experiences and other mental events individually have metaphysical subjects. I'll be questioning a stronger assumption that is often made: that not only do mental events and states in a single conscious life each have a metaphysical subject, but also they have the *same* subject. For example, if you are simultaneously feeling a tingling pain and hearing a tingling bell, we might naturally assume that these events have the same metaphysical subject—you! This motivates a distinction between *strong* and *weak* metaphysical subjects. Strong metaphysical subjects only exist if mental events in a single mental life have the same metaphysical subject; weak metaphysical subjects don't require this: they are simply the metaphysical subjects of experiences, whatever they happen to be. I will defend the view that there are *no strong metaphysical subjects*. More specifically, I will defend the view that even experiences *happening at the same time* within a single conscious life need not have the same metaphysical subject. For example, it might be that a visual experience has as metaphysical subject a different brain area from a simultaneous auditory experience, even though they are unified into a single conscious field. (Note: this is not the same thing as saying that brain areas "have experiences": see section 2 below, p.10). I'll call the view that there is such non-identity of metaphysical subjects the *Subject Non-Identity Thesis*.

To make the thesis vivid, let me contrast two different forms that a psychophysical theory might take. On the traditional subject-centered view, to explain the emergence of a particular stream of consciousness, we must first specify the subject of the relevant experiences (usually supposed to coincide with a whole organism). A complete description of their stream of consciousness would then involve specifying the phenomenal properties they enjoy at different times. These phenomenal events would then be explained in terms of a series of physical states *of the subject* that realize each phenomenal event. (For this to be explanatory, we would presumably also have to have a theory of how this realization is possible, given in terms of systematic connections between the relevant properties that have in some way been made intelligible to us).

On the less familiar subject non-identity view, different phenomenal events in a stream of consciousness (including those happening at the same time) need not have the same metaphysical subject. For example, their metaphysical subjects might be different regions of space-time (such as different regions of a subject's brain at different times), or different fusions of fundamental entities like particles occupying different regions. Each such metaphysical subject will have physical properties that realize the relevant experiential properties. If you consider the

4D region that is occupied by a person during her life, there will be very many 4D sub-regions of the worm occupied by experiences, which will be connected together in various ways to realize a unified stream of consciousness, which on this view is most perspicuously thought of as happening *in* the person, rather than *to* the person[6]. The subject's total experience at a time, and experience over time, are thus most perspicuously thought of as plural events, involving the subjects of the different experiential parts (e.g. different brain areas) as actors[7].

To sum up: events, including experiences, have metaphysical subjects. It is typically assumed that the parts of your experiences at a time have a single metaphysical subject; the subject non-identity theorist denies this. Below I give a more precise formulation of the thesis. First, I want to discuss some important epistemological issues that the notion of a metaphysical subject raises.

2. Identifying the Metaphysical Subject of an Experience

To decide between the competing pictures just described, we need to figure out what the metaphysical subjects of different experiences in a stream of consciousness are. In particular, we would like to know certain objective features of them like their spatial boundaries. Now, it is often simply assumed that the subjects of experiences are persons, which are usually taken to be coincident with whole human bodies. This can make it seem like there is no interesting issue about the identity of the subject of experience. I think this is mistake: this is a substantive empirical question. Moreover, there are some features of the situation that make determining this subject quite problematic: I call this the *self-predicament*. Before proceeding we need to understand why this is.

The reason it is an empirical question what the metaphysical subject of an experience is, is that we are able to pick out our experiences through introspection in a way that doesn't require us to attribute any objective physical properties to their metaphysical subject[8]. For example, we can coherently entertain the hypothesis that the external world doesn't exist, and that our experiences are the only things that exist. Or we can entertain the hypothesis that the subject of experience is a non-physical object like a soul. So, in particular, it is not part of my introspective concept of my current experiences that they have a human-body shaped object as metaphysical subject.

Another way of looking at this is that we can pick out a metaphysical subject of experience by description, simply as the thing that instantiates a certain experiential property, where our concept of the experiential property is based on something like introspective acquaintance; (I won't attempt to say more here about what that is). Even if this experiential property is identical with, or realized by, a physical property of the subject, our introspective concept of it doesn't reveal what this physical property is. (This is one aspect of the "inferential gap" that many philosophers believe divides physical facts and the phenomenal facts). Neither does our introspective concept of the experiential property reveal any

other objective features of the experience (such as where it is located in space, or what shape it is) that would give us identifying knowledge of its subject[9].

So it's not knowable a priori what objective physical features the metaphysical subject has. Neither are these features revealed by any kind of direct observation. Even if it's true that we're directly aware of ourselves as embodied agents moving around in the world, our awareness of our bodies doesn't reveal to us whether or not they are (or are coincident with) the metaphysical subject of our experiences. Even if I suppose that I am a body-shaped physical object (and not, say, an immaterial soul), and that I am directly aware of connections between my experiences and my body, it's not obvious that I "have" my experiences in the sense that I am their metaphysical subject. Rather, having an experience might be analogous to having a body-part. I might "have" a pain only in the sense that the pain is going on *in* me: its logical subject might, for example, be a proper part of me.

Incidentally, this shows what would be wrong with arguing as follows:

(1) People have experiences.
(2) Therefore, people are the metaphysical subjects of experiences.
(3) People are mereologically coincident with whole organisms.
(4) Therefore, the metaphysical subjects of experiences are coincident with whole organisms.

The trouble with this argument is (2). It's not obvious that the "having" relation in (1) is the relation of "being the metaphysical subject of". Maybe the real metaphysical subject of the experience is a body-part like a brain part, and people only "have" experiences in the sense that the experiences happen *in* them, or are integrated within them in a particular way (e.g. by being integrated in their mind). Again, compare this with having a hand. Conversely, if I say that a brain area is a metaphysical subject of an experience, this is not the same thing as saying that the brain area is having an experience. That might be a bit like saying that my hand has a hand. Or to give another analogy, take the case of a neural firing. I might be able to meaningfully talk of a person "having" a neural firing, but clearly the person is not the metaphysical subject of the firing, the neuron is, because it's the thing firing (the person is the subject of "having a firing"). Conversely, it would be wrong to say that the neuron *has* a neural firing; there is a difference between firing and having a firing. In sum, uncertainty about what kind of metaphysical subject experiences have means that there is no theory-neutral agreement about what it is to "have" an experience. (Compare how there is also no theory neutral agreement about what "having" a body part is: is my hand literally part of me, or rather part of the body which I control, as a dualist might believe?).

Even if the metaphysical subject of a pain is not a whole organism, the whole organism can correctly be said to "have" the pain. But then there is a property of "having a pain" that the whole organism can be said to instantiate, and therefore

an event of the whole organism having a pain that can be said to occur. What would be wrong with identifying the pain with the event of the organism having a pain? Isn't it more plausible that this is the event that I refer to when I say "I'm feeling a pain" than some event that involves a mere part of my body? Following on from this, we might say that at best the relevant property that belongs to my brain-region is "proto-pain"—an important ingredient in pain, but not pain itself.

I think our conception of a pain is *consistent* with this view, but again, the matter is not an a priori one to be decided based on linguistic considerations. To see how there is a reasonable alternative, consider an analogy between experiential properties and colors. We typically ascribe colors to whole objects: for example, I might say "that car is red". Still, it is plausible that the colors of *surfaces* are really more fundamental: what makes a car red is that most of its surface is red. It would be strange to think of surface color as really "proto-color", rather than color in the full-blown sense, even if ordinary language color terms almost always predicate whole-object color. I think the reason is that we understand what color is through experience, and the color properties we perceptually experience belong primarily to surfaces, and only derivatively to whole objects. Similarly, we identify our experiences by *introspecting* them. As just noted, this doesn't require conceiving of the metaphysical subject of the relevant experiential properties as having any particular spatio-temporal boundaries. So picking experiential properties out introspectively, we leave it open that they might belong to brain-parts. This seems correct, even if in natural language, when we ascribe experiences to ourselves and others, we are always talking about properties of whole organisms.

If we can't know it a priori or by direct observation, how do we figure out what the metaphysical subject of experience is? Arguably, the only way to do this will be to find out which physical properties realize, or at least are correlated with, experiential properties[10]. Then we can figure out which object has the physical properties that are correlated with the experiential property we are interested in. So to identify the metaphysical subjects of experiences, we need to find the physical correlates of experiential properties without even first knowing *which* physical objects have these properties (that is, without knowing anything about what physical properties they have). This is the *self-predicament*.

This lack of identifying information about the metaphysical subject of experience is somewhat unique, in that typically when we observe some event, we get objective information about the identity of the metaphysical subject involved, such as information about it's spatial features. For example, if I directly observe an instance of walking, I typically observe the thing doing the walking as having certain spatial boundaries and other objective features (it might even be *inconceivable* that the walker fail to have certain objective features; for example, it is not conceivable that it is really my brain that walks). So if we want a theory of what walking is—or how it is physically realized—we do not need to figure out simultaneously which things are doing the walking. That we *are* in this

situation with our own experiences is part of what makes the problem of the self so interesting. (Note: if you think that a similar problem arises in the case of other events, that's ok with me: my concern is to show that there is a substantive epistemic problem identifying what the self is, not that there *isn't* a similar issue elsewhere).

We should not conclude that it is *impossible* to identify the metaphysical subject of experience, even if we assume that there is an inferential gap between the physical and phenomenal facts. Or at least, the self predicament may at most complicate the already difficult problem of identifying the physical properties underwriting experience, rather than making it insuperable. Even if there is a physical/phenomenal inferential gap, we may be able to discover these underlying physical properties empirically, for example, by exploiting our introspective knowledge of the content, structure and functional role of experience, and finding physical properties that have a corresponding form and role (although see Chalmers (1998) for an argument that there are fundamental epistemological limitations here). There is no reason in principle why we can't "solve simultaneously" for the metaphysical subject of experience and the physical correlate of an experiential property[11], even if in practice this is quite difficult.

What this should bring out, I hope, is that finding the metaphysical subject of an experience is a difficult empirical issue. Moreover, there is no particular justification for taking it as a working hypothesis that the metaphysical subject is coincident with a whole organism, as many authors do. In advance of investigation, we know very little about what the metaphysical subject is. It may even be that simultaneous experiences within a single stream of consciousness don't have the same metaphysical subject: the thesis I will now consider in detail.

3. A more precise formulation of the thesis

Bearing all this in mind, here is a first attempt at formulating the Subject Non-identity thesis slightly more precisely:

> **Subject Non-identity thesis** (first pass): Experiences occurring within a single unified conscious life have different metaphysical subjects.

This is probably too weak, given that 4-dimensionalism might be true. In that case, the metaphysical subjects of experiences happening at different times within a single unified conscious life will be distinct short-lived temporal slices. The Subject Non-identity thesis is supposed to be the more radical claim that even *simultaneous* experiences belonging to a single person can have different metaphysical subjects:

> **Subject Non-identity thesis** (second pass): Experiences occurring simultaneously within a single unified conscious life can have different metaphysical subjects.

Less obviously, even this is too weak. What does it mean to say that two experiences are *simultaneous*? I think that experiential properties are probably instantiated not by instantaneous temporal slices, but by slices that have some temporal breadth, because the minimal amount of neural activity required for conscious experience will involve extended processes like neural firings (see Lee (2014a)). This suggests that if 4-dimensionalism is true, "simultaneous experiences" should be understood as experiences whose metaphysical subjects are temporal slices that exist in *overlapping* temporal intervals. Insisting that they occupy the *same* interval is presumably too strong—for example, auditory experiences might invariably be realized over longer intervals than visual experiences. Thus, simply granting 4-dimensionalism and some plausible assumptions about experiential realization is likely to imply that our second version of subject non-identity is true. But I had something stronger in mind. I think we can capture it by appealing to the notion of "weak spatial coincidence": say that two objects weakly spatially coincide, just if at every time at which they both exist, they share all their spatial parts. I think the following formulation now captures the intuitive idea we are after:

> **Subject Non-identity Thesis** (Third pass): The metaphysical subjects of simultaneous experiences occurring within a single conscious life may not be weakly spatially coincident objects.

If this is true, then the metaphysical subjects of simultaneous experiences might not coincide, not only because they might not share all temporal parts, but also because they might simultaneously have different spatial parts. This version of the thesis captures the idea I want to explore here. In the next section I turn to considering the positive case that can be made for it, before responding to objections in section 5.

4. Arguments for the Subject Non-identity Thesis

In this section, I consider three arguments that might be given for the Subject Non-identity thesis. The first argument is only supposed to show that there should not be a default presumption that the Subject Non-identity thesis is false; it therefore only plays a supporting role. The second argument doesn't succeed on its own in eliminating alternatives to the thesis, but it is still illuminating to briefly consider it, in part because it helps motivate the third argument—the Subtraction argument—which I take to provide the strongest case for the thesis.

4.1 The 4-Dimensionalist Argument from Analogy

We noted above that if 4-Dimensionalism is true, it is plausible that experiential properties are not instantiated most fundamentally by instantaneous slices, but rather by temporally extended slices. This is plausible because experiential

properties are probably realized by extended physical processes like neural firings. If this is correct, then on a 4D view experiences within a conscious life happening at the same time probably fail to have the same metaphysical subject for the simple reason that their metaphysical subjects have different *temporal breadth*—for example, perhaps typically auditory properties are physically realized over longer intervals than visual properties. So, on a 4D view, the closest we might hope to get to strong metaphysical subjects are metaphysical subjects that are *weakly spatially coincident*: they share all spatial parts whenever they exist at the same time. However, weak spatial coincidence is not identity, so in conceding this, we have already given up on the intuitive idea that simultaneous co-streamal experiences have the *same* metaphysical subject.

A 4-dimensionalist might now argue as follows: once we have given up on the identity of metaphysical subjects of simultaneous unified experiences on the grounds that they might not share all *temporal* parts, why think that they must share all their *spatial* parts also? Why think that there is a disanalogy between time and space here? Raising this challenge doesn't show that these metaphysical subjects are *not* weakly spatially coincident, but it puts the proponent of strong metaphysical subjects on a back foot. If they can't answer the challenge, then at least we might conclude that their view does not deserve a default status.

4.2 The Duplication Argument

Another line of argument that is worth briefly discussing is the following (related arguments appear in Peacocke (1995) and Parfit (1999), although they are not explicitly concerned with the SNI thesis). Consider the fact that it is surely possible for two experiences of the same kind to occur within the body of a single organism. For example, a two-headed organism like the Pushmi-Pullyu (the two-headed gazelle encountered by Dr. Doolittle in Lofting (1998)) might have two separate brains capable of supporting qualitatively identical experiences. To understand this, we need to suppose that the organism's enjoyment of different experiential properties is in some way relativized to different parts of its body. For example, she can enjoy the color blue either relative to head 1 or head 2.

One possible analysis of this relativization would support the subject non-identity thesis. It might be that what it is for the organism to enjoy a certain kind of experience relative to a body part is for the body part to instantiate an experiential property in a more fundamental non-relative sense. This would support the subject non-identity thesis, at least provided we had some reason to think that in an ordinary person different experiences are relativized to different body parts.

The problem is that we need some reason for preferring this to another interpretation on which the thesis is not supported. To see the alternative, consider, by way of analogy, the property of waving. Since a person can wave with either left hand, right hand, or both, we might conclude that the correct metaphysics of

waving involves a relativization to a particular hand. But it would clearly be an error to conclude from this that is really the hand on its own that is waving, and *you* are only waving in the sense that one of your hands is waving. If a detached human hand shakes back and forth in space, that is not a waving. At best it is a "quasi-waving"; to be a real waving, it must be controlled in the right way by a person whose body it is attached to. Thus, despite hand-relativization, it is plausible that the whole organism is the metaphysical subject of the waving.

A similar interpretation could be given of experience-relativization: your brain-part is at best quasi-experiencing, not really experiencing. Thus, to defend the subject non-identity thesis, we need to establish that experiences are disanalogous to wavings in this respect. We need to argue that not only is there neural activity in your brain that is of special relevance to the existence of your experience (in the way that hand movement is of special relevance to waving), we need to establish that this neural activity *on its own* should plausibly count as an experience.

Fortunately, there is an argument that supports this interpretation, the Subtraction argument. Furthermore, if it works, it supports the Subject Non-identity thesis directly, rendering considerations of duplication superfluous[12].

4.3 The Subtraction Argument

The subtraction argument starts with the observation that most of the events happening in your body at a given time are at best relevant to what experiences you have only in a causal sense. The events that are constitutively relevant—i.e. they are part of a minimal realization of the experience (and here I have in mind the *total* realization (Shoemaker (2007)))—are, at least on many views, fairly localized; for example, many would say that they are localized to areas of the brain. If this is right, then in a situation where only the localized realizer events happen, and they are not embedded in a larger biological surround, they would still realize the existence of the same type of experience. It is inferred that even when the events occur in a biological surround, the surround is not an essential part of the experience. Therefore, the metaphysical subject of the experience must have its existence realized by the more localized events.

The core of this argument can be laid out as follows:

(1) Of the physical events occurring in a subject S's body during the time at which experience E is occurring, there is some minimal subset SUB of these events, such that were these events to occur in a detached form—i.e. without the bodily events that actually surround them—an experience of the same type as E would still occur. Call this counterfactual experience C(E).
(2) The metaphysical subject of C(E) is some proper part P of S's body.
(3) C(E) and E have the same metaphysical subject.
(4) Therefore, the metaphysical subject of E is P - a proper part of S's body.

The conclusion does not imply the subject non-identity thesis, but could be extended into an argument for it by giving an argument that the minimal realizers of different parts of a unified experience will not, in general, involve exactly the same body parts: for example, that auditory and visual experiences do not involve exactly the same brain areas. Although this is quite plausible, there is an important view incompatible with it, phenomenal holism, which I will discuss below.

(1) is the least controversial premise of the argument. Strictly speaking, (1) does not even state that SUB is a *proper* subset of bodily events (i.e., it might contain *all* the events occurring within the subject's body), it just states that SUB exists[13].

(2) only requires that there are *some* parts of the subjects body, such that SUB does not contain events involving these body parts. Presumably some such parts of S, such a leg-hairs, exist, so this is also extremely plausible. It does not require that the metaphysical subject of C(E) be a highly localized brain-part, merely that it is some proper part of S.[14] (Although again, to establish the subject non-identity thesis we will need to establish that the metaphysical subject C(E) is a *different* body part for different simultaneous experiences, and this will involve making more specific claims about *which* body parts are involved in SUB (they may well be brain-parts). The Subtraction argument, as stated, does not try to establish anything about this.)

Premise (3) is the controversial premise. Tye (2003), in a related discussion, points out that it is fallacy to infer from the fact that if a proper part of X existed in detached form it would be an F, to the conclusion that it is an F. For example, a part of a chair might be such that if it existed on its own it would be a chair. It doesn't follow that it actually is a chair. (As Sider (2001a) points out, this means that "chair" is not an intrinsic property (see below))[15]. One might worry that accepting premise 3 would involve a similar fallacy. Compare the case of my walking to my experience: even if there is some set of events going on in my body that on their own would by sufficient for a walking to exist, a walking whose metaphysical subject would be some proper part of my body as it actually is (e.g. me minus one leg hair), it doesn't follow that the my actual walking has as metaphysical subject some proper part of my body. So we should agree that if supporting premise (3) required an extra premise such as the following, then it would not be a good argument:

> **Subtraction**: If a group of events is such that, were they to exist in a detached form then that would be sufficient for an event consisting of object O X-ing to exist, then they are actually sufficient for an event consisting of object O X-ing to exist.

However, perhaps there is a more restricted principle that applies in the case of experiences, but not, say, in the case of walkings, which might allow us to defend (3). One obvious idea to try here is to stipulate that X be an *intrinsic* property, in the sense that something having it doesn't depend on how

the material that constitutes it is embedded in the world, only on how it's parts are (intrinsically) propertied and related to each other[16]. Then the modified principle is clearly true:

> **Modified Subtraction**: If a group of events is such that, were they to exist in a detached form then that would be sufficient for an event consisting of object O X- ing to exist, and X is an intrinsic property, then they are actually sufficient for an event consisting of object O X-ing to exist.

The modified principle doesn't apply to properties like being a chair, being a rock, or walking, because, as mentioned, they are *not* intrinsic properties, in the relevant sense; their application depends on how the object's parts are embedded in the world (Sider (2001a)). For example, roughly speaking, an object has to be detached to be walking, and being detached is an extrinsic property. If experiential properties are, by contrast, intrinsic, then we can complete the Subtraction argument using the modified subtraction principle.

So, are experiential properties intrinsic properties? Despite being an intuitive claim, and one I myself and plenty of other theorists would accept, this would not be accepted by all consciousness theorists, as I will explain. However, even if experiential properties aren't intrinsic, this is for quite different reasons from those associated with properties like walking or being a chair. As mentioned above, experiential properties might be extrinsic because they are relations to the *targets* of experience (such as external objects), whereas "being a chair" is extrinsic because it constrains the *spatial boundaries* of the chair. It will be helpful to spell out this difference, as it will enable us to see why either the modified subtraction principle is true, or a related principle that serves our purposes while allowing for certain ways in which experiential properties might be extrinsic, is true.

Which theories of experience reject intrinsicness? Two popular theories are salient here. First, externalist intentionalists (e.g. Lycan (2001), Tye (2002)) hold that perceptual experiences are states of sensory representation, representing the environment as being a particular way, and that the representation of features of external objects and events requires external links between the subject and these features, such having internal states that reliably detect these features. Second, acquaintance theorists (e.g. Campbell (2002), Martin (2003)) hold that perceptual experiences involve the subject of experience standing in a relation of acquaintance to external objects and events; these experiences therefore require that these external objects exist.

So, experiential properties might be relational, because they are relations to the *targets* of experience (object, properties, events). By contrast, the reason that a property like "being a chair" is relational is that it is *spatial boundary constraining*, in the sense that: (1) a chair must have a *natural* spatial boundary, and (2) the natural boundary must have certain spatial features (e.g. it allows the object to be sat on, given the material it is made from). Most of the mundane macroscopic objects we are interested in have a natural spatial boundary, in the

sense of a boundary where there is a change of material, e.g. from a gas or liquid like air or water, to a more solid material like skin, fur, wood, metal or plastic. We are often interested in the shape of this natural boundary, and the way it changes over time (e.g. this is part of what we are describing when we say that a person is walking). In fact, it probably isn't an exaggeration to say that *most* of the properties of material objects we normally talk about involve the shape of the object's natural boundary, and therefore require that the object *have* a natural boundary (so they are extrinsic, requiring the object to be surrounded by a different material).

Arguably, experiential properties are not boundary-constraining in the same way, even if they are extrinsic. We pick out experiential properties *introspectively*, and do not conceptualize them in terms of their metaphysical subject having natural spatial boundaries, let alone natural spatial boundaries that have a certain shape. For example, one of the reasons why scenarios like *the brain-in-a-vat*, or *disembodied consciousness*, are at least conceivable is that our experiential concepts are not boundary constraining spatio-temporal concepts. Compare how disembodied chairs are not conceivable, because it is a priori that chairs have natural boundaries. (This lack of a priori boundary constraints is part of what creates the self-predicament mentioned earlier; it is one way in which we don't know much in advance about what kind of object the metaphysical subject of experience is).

If experiential properties aren't boundary constraining (I'll defend this further in a minute), then the following principle, which can play a similar role to modified subtraction, is plausible:

> **Boundary Addition Principle**: Adding extra material (such as the rest of a subject's body), around the spatial boundary of the metaphysical subject of an experience (such as C(E)) doesn't prevent that object from having experiential properties, provided relevant external relations (such as those required for representation or acquaintance) between the object and the environment remain in place.

To elaborate: in our counterfactual world, a proper part of the subject's body (call it PART), such as a brain-part or larger body-part, exists in a detached form, in the same physical state that it is in the actual world. PART's state realizes the same type of experience (call it EXP) in both cases (although we cannot assume PART, rather than a larger object, instantiates EXP in the actual scenario). Also, if EXP requires external relations to the environment, then these relations hold between PART and the environment in our counterfactual world, and also in the actual world, where EXP is still instantiated, even if not by PART itself. The difference between the scenarios is that in the actual world, PART is surrounded by additional material (the rest of the subject's body). This *does* change PART's relational properties, but if EXP (unlike "walking" or "being a chair") isn't spatial-boundary-constraining, there is no reason to think this surrounding material stops PART from instantiating EXP, since by stipulation the *relevant* external relations between PART and the environment are not changed

by adding the subject's body back into play. So its plausible that PART *does* instantiate EXP in the actual scenario. (Compare how a relational property like walking to the store *doesn't* satisfy an analogue of boundary addition, because walking *is* a spatial-boundary constraining property).

Thus, we can give a sub-argument for premise (3). It looks like this:

(S1) Experiential properties are not spatial-boundary-constraining.

(S2) If they aren't spatial-boundary-constraining, then the boundary addition principle holds.

(S3) If the boundary addition principle holds, then E and C(E) have the same metaphysical subject.

I think the most contentious step of this argument is probably (S1), so let me say some more to defend it. It might be objected that, even if there aren't natural-boundary constraints explicitly built into experiential concepts (so, e.g., disembodied minds are conceivable) this doesn't imply that they do not pick out properties that *in fact* involve such constraints, even if it does eliminate one important reason we might believe they do.

I think we can't decisively reply to this objection without knowing what the correct theory of consciousness and experiential properties is, and this is something that is completely up in the air. However, we can at least show that (S1) is plausible over a wide range of different views, and therefore that it should be attractive to many theorists.

One important preliminary point here, which many theorists will agree with, is that a particular experience-type can usually be associated[17] with objects that have wildly different spatial boundaries. Suppose, for example, that a brain-in-a-vat scenario is not only conceivable, it is also possible. That is, suppose it is possible for experiences of the type you are currently having to be realized in the physical states of a disembodied brain. A brain has very different spatial boundaries from a human body. It would be surprising if one and the same experiential property constrained its subject to have a human-body-shaped natural boundary in one case, and a brain-shaped natural boundary in another case. A more plausible explanation for such variety is that, unlike "being a chair", having such a property does not place any spatial-boundary constraints on its metaphysical subject or associated objects. Compare walking: there might be some part of my body (e.g. me minus 1 leg hair) such that were it to exist in a detached form, it would be walking. But it would have a similarly shaped natural boundary to me, and that's because walking *is* spatial-boundary-constraining.

More significantly, S1 (and the above explanation of shape-variety) is supported on a wide range of specific theories of what it is to have an experience. Three of the most popular theories are representational theories, acquaintance theories, and qualia theories; they all support the view that experiential properties are not spatial-boundary-constraining.

As I mentioned above, on representational theories (e.g., Dretske (1997), Lycan (1996), Tye (2002)), experiential properties are a special kind of intentional property; an experience "sensorily represents" the world to be a certain way. The physical basis for these sensory representations is typically thought to involve the subject's brain being in a representational state (perhaps with a certain internal representational structure), a state which gets its intentional properties and counts as a conscious sensory state in virtue of playing a certain functional/computational role, and being hooked up to the environment in the right way. Plausibly, both a disembodied brain and a normal human body could be associated with such representational state, a reflection of the fact that it does not seem to have anything to do with the object's spatial boundaries[18]. An analogy might be with a computer running a particular software program—this tells us nothing about what shape the computer or any of its parts have, and doesn't even require that the thing doing the computing (which may be some part of the computer) have natural boundaries.

Second, on acquaintance theories, experience involves the subject being acquainted with external objects and events (e.g. Campbell (2002), Martin (2003)), or with more exotic objects like non-physical sense-data (e.g. Russell (1912), Jackson (1977)). Prima facie, being acquainted with something has nothing to do with one's spatial boundaries, and in particular does not seem to require natural boundaries. Rather, if it has some kind of analysis in physical or functional terms, it would seem to have to do with an informational channel (e.g. involving light waves bouncing off an object and causing information processing in the brain) existing, linking the metaphysical subject of experience to external events. There appears to be nothing to prevent a brain or brain-part (rather than a whole organism) standing in such a relation to the environment, again reflecting an apparent lack of boundary-constraint. Alternatively, acquaintance might be a primitive relation with no natural analysis; but again there seems to be no reason why such a primitive relation would impose constraints on the spatial boundaries of its subject.

Third, qualia theories (e.g. Block (2003), Papineau (forthcoming)) say that experiential properties do not relate the subject to some external entity like a content, sense-datum or external event, but are rather monadic sensational properties (they may nonetheless have a rich structure). Again, nothing about this seems to have anything to do with the subject's spatial boundaries; typically, it is thought to involve having an *internal* state of a certain kind, where "internal" connotes that the relevant property has to do with the internal organization of the experiencer, rather than the experiencer's spatial boundaries.

Admittedly, these remarks may be too brief to convince everyone. Furthermore, there are other theories of experience, or ways of elaborating the above theories, that may imply that the metaphysical subject of experience has a natural spatial boundary with certain properties (e.g. it is human-body shaped). For example, according to some versions of "embodied" theories of experience, experiential properties involve a complex feedback process between the environment

and the subject's body, and this may require that the body exist (although having said that, there are surely versions of the view on which relevant feedback processes involve different *parts* of the body rather than the whole body). A full defense of the subtraction argument would have to rule out these views, and to look more carefully at the views discussed: but that is beyond the scope of this paper. For now, I hope to have done enough to persuade many theorists that S1 is plausible.[19]

This concludes my defense of the Subtraction argument. Now, as mentioned earlier, the Subtraction Argument does not take us the whole way to the subject non-identity thesis, because even if the metaphysical subject of an experience I'm having is a proper part of my body, this doesn't imply that it is ever a *different* body part for different experiences. When we consider different kinds of experiences, such as visual and auditory experiences, it is quite plausible that their minimal neural bases are not wholly coincident (although they may greatly overlap). If the subtraction argument works, this suggests that they have different metaphysical subjects. Having said this, there is a general view of the structure of experience on which different parts of experience will have the *same* metaphysical subject—*phenomenal holism*. To complete my defense of the Subject Non-Identity Thesis, I need to explain why I reject this view. (Elsewhere I discuss the view in much more detail (Lee (2014b)).

4.4 Phenomenal Holism

According to the phenomenal holist, the most basic units of conscious experience are total phenomenal fields. Experiences do have parts, but these parts are mere abstractions from the whole, owing their existence to the whole field. If we think of experiences as property instantiations, then we can look at it like this: experiences that are "parts" of larger experiences are really instantiations of properties that are derived from more fundamental total phenomenal properties (the relationship can be thought of as at least analogous to the relationship between a determinate property and the determinables that it falls under). An analogy would be with a holistic metaphysics of fundamental reality (e.g. of the kind endorsed by Schaffer (2010)), on which local facts are all derived from the global state of the whole universe. Such a view is sometimes taken to be implied by an interpretation of quantum mechanics on which the state of the global wave-function is the basis for everything else.

Note that phenomenal holism implies the existence of a momentary strong metaphysical subject, because the total experiential property will be instantiated by a certain object, and derivative phenomenal properties—corresponding to "parts" of the experience—will be instantiated by the *same* object.

How are we to assess this view? If we are physicalists, then we might think that experiences have a total physical realization: a set of physical events that are minimally sufficient for the existence of the experience. It is a tricky matter how

exactly to elaborate this notion of "realization", but I think that it is plausible to understand it in such a way that if phenomenal holism is true, then each part of an experience has the same total physical realization (see Lee (2014b)). If this is right, then phenomenal holism and "realization physicalism" entail:

> **Total Realization Holism**: Different parts of a total experience have the same total realization.

Clearly, this is incompatible with the subject non-identity thesis. How plausible is it? As already mentioned, even if there is a lot of overlap in the total realizations of different parts of a conscious experience, it is not especially plausible that they must totally overlap. For example, the total realization of a conscious visual experience will involve events in the visual system that prima facie have nothing to do with the realization of a simultaneous auditory experience, whose realization will include events in auditory areas, and other areas, but not events in the visual system[20].

Admittedly, I can imagine views on which, despite initial appearances, this is wrong. Suppose we hold a representational theory on which conscious events consist in mental representations having a certain separable property of "being conscious". For example, on a basic version of the higher-order view, a conscious mental state is a representation that has a higher-order representation directed on it, representing the fact that the first-order state exists (as on the earlier versions of Rosenthal's view (e.g. Rosenthal (1990)); or on Prinz's (2012) attentional view, a conscious representation is one that is attended, where "attention" is conceived of as mechanism that makes representations available to working memory. We can imagine views in this framework on which "being conscious" is a holistic property that depends on what is happening over the whole system. For example, imagine a model on which the neuronal populations corresponding to different representations competing for consciousness all have a certain activation level, and what determines which representation is conscious at a given moment is simply which has the highest activation level (so only one is conscious at a time). Being conscious will be a holistic property of the representation, because we can't tell whether a given representation is conscious without first checking that there aren't any other representations in the system that have a higher level of activation. Similarly, even if more than one state is conscious at a time, on a more realistic model it could still be the case that "being conscious" is a similarly holistic property of a representation; that is, the fact that it is conscious is not separable from the facts about what is happening with other representations, both conscious and unconscious in different parts of the brain.[21]

However, I am doubtful whether "being conscious" is holistic in this way. We are still grappling around in the dark when it comes to understanding what consciousness is, so it is hard to say anything with confidence in this area. But certainly existing theories are probably not best understood as holistic in this way. The theories that probably have the best chance of being plausibly

developed in a holistic way are those that take the consciousness of modality specific representations as dependent on some shared central resource, such as working memory, attention, a global workspace, or a faculty of meta-psychological representation. For example, if there is some limit to the number of representations that can have the resource applied to them, this might be thought to give rise to a holism of the relevant kind.

However, I suspect that on most plausible views of this kind, the only holism is causal rather than constitutive, in the following sense. If representations are competing for some limited central resource, then the process whereby they are selected may involve complicated feedback across the brain between different regions—processes which might be described as "causally holistic". But if a representation becomes conscious as a result of such a global process, it may well be that it is conscious in a constitutively non-holistic way. To give a silly analogy, imagine an army of zombies fighting to get through the gates into the citadel of consciousness; only a small number of zombies can fit through the gates into the citadel. The process whereby zombies get selected may involve complex causal interactions across the whole army; but once a zombie gets selected, the fact that is in the citadel may be a more local fact. I suspect that viable central-resource theories are most likely to have this non-holistic form, although I can't say with great confidence that this is correct.

To sum up: more could be said about phenomenal holism than I have space for here. But given the prima facie plausibility of the claim that modality specific experiences have total realizations that don't completely coincide, it seems to me that the phenomenal holist, if she believes that experience is physically realized, needs to work hard to justify their view: the default should be that they are wrong. If phenomenal holism fails because of non-coincident total realizations, and if the subtraction argument is good, then we have good reason for thinking that different parts of a total experience have non-coincident metaphysical subjects. For example, the metaphysical subject of my visual experiences may includes parts of the visual system that are not parts of the metaphysical subject of my auditory experiences. That is, we have good reason for thinking that the subject non-identity thesis is true.

This concludes my positive defense of the main thesis of the paper. In the next section I respond to a number of important objections to the thesis.

5. Arguments against the Subject Non-identity Thesis

5.1 A "No Self" view is implausible

I take it that the metaphysical subject role is a central strand in the conception of the self that many theorists of the self have in mind. When people say that the self is the *subject of conscious experience*, this is a good candidate for what they have in mind. Moreover, it is generally assumed that different parts of

your experience have the same metaphysical subject. If this is wrong, then in an important sense, the "subject of experience" *doesn't exist*.

This might seem hard to accept. Can't I know with something close to certainty that I exist? In response, I think a large part of any seeming craziness of the view can be defused by pointing out that even if there are no strong metaphysical subjects, there still might be "selves" in other senses. There are a number of other important self-roles that are also common strands in the self-conceptions that theorists have in mind. In my view, they are not a priori equivalent, and so cannot be assumed to pick out the same object. Of special importance here is the point that, if the subject non-identity thesis is correct, then these other "selves" probably still exist, although they are probably not identical to the (weak) metaphysical subjects of any experiences. This latter point will help us understand the broader significance of the thesis, as I will explain.

What other self-roles are important? One is of the self as object of self-reference. Both in thought and public language use we refer to ourselves using the first-person pronoun. In recent discussions of the self, particularly those in the neo-Kantian tradition (starting with Strawson (1966)), it has been common to shift focus away from first-order questions about the existence and nature of the self, to questions about first-personal thought and reference. Such theorists seem to be assuming that "the self", whatever it is, is most helpfully thought of as just whatever the reference of "I" is, and so we need to understand the nature of self-reference to understand the self[22].

A second important self-conception, related to the last one, is the self as object of self-awareness. Instances of self-awareness perhaps do not always involve a fully conceptual use of the first-person concept, and for this reason will, at least on some views, involve a different category of mental states from first-personal thoughts. An example worth mentioning here are views on which the "subjectivity" associated with phenomenally conscious states is to be understood in terms of self-awareness, as on the "same-order" view of consciousness (which may go as far back as Aristotle (Caston (2002)); see Kriegel (2009) for a modern development). On this view, all conscious states include in themselves an awareness that their subject is in the state in question. If it is true that a conscious life is thoroughly permeated with self-awareness in this way, we can ask which thing, if any, is the "self" that is the object of the awareness[23]. Another form of self-awareness that is relevant is the self-awareness that comes with the use of egocentric frames of reference in perception: for example, visual awareness uses a head-centered spatial frame to represent the positions of objects around the perceiver. It is probably true that at least some of the sense we have of the self as the "origin of the world" comes from the use of egocentric frames such as this.

A third self-conception we might be interested in the agential self: the self as the originator of, and subject of, physical and mental actions. Finally, it has been common to see the self as having a special role to play explaining the unity of consciousness (although it is notoriously unclear what this unity consists in—see Bayne and Chalmers (2003) for a helpful taxonomy of different notions of

"unity"). Simply belonging to a single self is sometimes thought to explain unity; another common view is that being jointly accessible or introspected by a single subject is what constitutes unity.

In each case, if there is no strong metaphysical subject, it is plausible that the role is still satisfied, but that it is *not* played by a (weak) metaphysical subject of experience. Start with the self as object of self-reference. If I think "I am in pain; I am hearing music; therefore, I am feeling pain and hearing music", its plausible that the different uses of "I" are semantically linked, and so refer to the same object. This remains true even if the experiences have different metaphysical subjects. In that case, it is much more plausible to say that each use of "I" refers to a single object that "has" all these experiences, such as a human organism (note that the "having" relation here won't be the relation of being metaphysical subject), rather than saying it fails to refer at all. So even if there is no strong metaphysical subject, you can still truly think "I exist".

A similar point seems plausible for non-conceptual self-awareness. If consciousness involves Kriegel-style self-awareness, then it is plausible that different instances in a single stream (especially if they are simultaneous or in very close temporal proximity) are semantically linked in such a way that refer to the same "self", which may not be the metaphysical subject of experience. In the case of egocentric content in perceptual experience, it is plausible that the "self" that is at the origin of the perceptual frame is the subject's body or part of the body, like the head. This may well be a different object from a (weak) metaphysical subject of experience, which on the subject non-identity thesis might be an internal body-part like a brain-part. (Note that even though self-awareness is, by definition, directed at the thing that is self-aware, we should not automatically conclude that it is directed at the metaphysical subject of the self-awareness. A thing might "have" self-awareness in the way required for the self-awareness to be directed *at* them, even though the relevant "having" relation is not the relation of "being the metaphysical subject of". (See remarks above on the "having" relation.))

In the case of the agential self, if there is some object that can be singled out as the most fundamental subject of (or originator of) physical and mental actions (and I do not say this is a very clearly defined role), it is presumably either the whole human organism, or perhaps some part of the organism that realizes the subject's mental states, in particular those responsible for the initiation of action, such as desires and intentions. This object could still exist even if there is no strong metaphysical subject. Furthermore, if different parts of your experience at a time have different metaphysical subjects, it seems unlikely that any of them are exactly coincident with either the whole organism, or the brain-parts that initiate action in the relevant sense (e.g. your desires are presumably not even partially realized in the parts of visual cortex relevant to visual experience).

Finally, when it comes to the unifying self, there may be interesting self-roles defined in terms of experiential unity that are satisfied even if there are no strong

metaphysical subjects. One particularly salient notion of this kind is that of a "stream-subject"—a fusion of all weak metaphysical subjects associated with the components of a unified stream of consciousness. This is an object that occupies the region of space-time where the stream of consciousness takes place. There is a natural sense in which this object is the "subject" of the whole stream of consciousness. If the subject non-identity thesis is correct, it is not identical with the metaphysical subject of a particular experience, but rather has each such subject as a part; so again, the self-roles pick out different objects. I won't look at other ways the self could be understood in terms of experiential unity, but I suspect a similar conclusion would hold for them.

So, even if there is no self *qua* strong metaphysical subject, there are probably these other 'selves', which furthermore are probably *not* coincident with the weak metaphysical subjects of experiences if the subject non-identity thesis is true. So the subject non-identity thesis is not a "no self" view in an unacceptably crazy sense; for example, it doesn't imply that you don't exist.

You might be tempted to think that this same point detracts from the significance of the thesis: it only goes to show that other self-roles are more important in understanding the self than the metaphysical subject role. But as I suggested earlier, this would be a mistake. The thesis would still be highly significant, even if we were to conceive of the "self" entirely in terms of self-roles other than the metaphysical subject role (which surely *is* a central strand in many self-conceptions). The reason is that all the "selves" mentioned here plausibly exist at least partly in virtue of the existence of a conscious mental life (or perhaps a temporal part of it). On the other hand, if the "selves" in question are not the metaphysical subjects of experiences, this suggests that these experiences will not be individuated in terms of these "selves", and so these experiences do not depend on these "selves" for their existence (this claim does need some qualification—see endnote)[24]. This suggests a kind of moderate bundle theory, in the sense that there is an asymmetrical dependence between selves and experiences: the "self" is partly constituted by experiences whose identity does not involve the relevant "self".

Below (section 6) I say a little more about what is involved in holding such a bundle theory. The connection between different versions of the bundle theory, the "no self" view, and the subject non-identity thesis certainly deserves extended discussion, which I won't provide here. For now, I hope to have given a clearer sense of how the different self-roles relate to one another, and the possible significance of the subject non-identity thesis. Let's turn to some other objections to the subject non-identity thesis.

5.2 The Cartesian Inference Objection

If there is no guarantee that "I" refers to the metaphysical subject of my experiences, then this raises the worry that my experiences existing do not guarantee

that I exist. Consider, for example, the following inference:

(1) This experience (i.e. this experience that I am aware of introspectively) is an experience as of a green object.
(2) Therefore, I am having an experience as of a green object.

Call this a "Cartesian" inference (in one more step, one could infer "I exist"). It's plausible that this is an a priori inference (i.e. the conditional "if (1) then (2)" is a priori). If the metaphysical subject of a given experience is not a priori identical to the person referred to by "I", does this threaten the a priority of the inference? If so, is that a problem?

There seem to be only two ways in which one could challenge the move from (1) to (2). First, one could challenge whether the person having the experience is the same as the person referred to by "I". Second, one could challenge whether "I" succeeds in referring to anything at all. The second challenge is more relevant here[25]. A proponent of the challenge will ask: is it a priori that "if this experience exists, then I exist"? Call this the "Lichtenbergian challenge" (Lichtenberg famously questioned whether this conditional (or something similar to it) is a priori (Lichtenberg (1971)). Acknowledging the possible truth of the subject non-identity thesis does affect how we should think about this. If it were a priori that "I" refers to the metaphysical subject of current experiences, then it is clear that the conditional *would* be a priori (assuming it's a priori that an experience has a metaphysical subject). Someone who takes seriously the idea that the metaphysical subject of an experience might be a mere part of themselves, however, won't accept this claim about "I", and so one possible defense of Descartes against Lichtenberg isn't open to them.

However, that doesn't mean that the Cartesian conditional isn't a priori. It might be a priori that a current use of "I" in thought refers to whoever "has" your current experiences, and a priori that someone "has" the experience, even if it isn't a priori that this is the metaphysical subject of experience, or a priori what exactly it takes to "have" an experience. For example, perhaps I can know a priori that *if* the metaphysical subject of this experience is embedded in a larger system like a human body, then "I" refers to this body-coincident object, but if it *isn't* embedded in such a larger system, then "I" refers to the metaphysical subject itself. Assuming these possibilities cover everything epistemically possible (e.g. there's no possibility where this experience lacks *any* metaphysical subject), then I know a priori that "I" refers, and so I exist, and the Cartesian conditional is a priori.

Such an a priori connection between my experiences and my existence is consistent with my experiences failing to guarantee I exist, in a certain sense. The important point here is that even if the metaphysical subject of this experience is a mere part of me, and this part (and hence the experience) *could have* existed without me existing (counterfactual possibility), and so my experience doesn't guarantee in a *modal* or *counterfactual* sense that I exist, it doesn't follow that there is an *epistemic* possibility where this experience is not my experience, or where I don't exist.

Having said all this, I don't think it would be unreasonable to side with Lichtenberg instead: maybe there really are epistemic possibilities in which you don't exist. It's not unreasonable to think that an experience is only had *by* someone if it is embedded in some larger psychological system. If it is not a priori that some such system exists (or doesn't follow a priori from the existence of the experience), then maybe the Cartesian conditional isn't a priori. Lichtenberg wins. It's true that allowing for the epistemic possibility that you are not identical with the metaphysical subject of your experiences makes this Lichtenbergian view more viable, but that doesn't in itself seem objectionably counterintuitive to me.

5.3 The "What it's like for me" objection

Another objection appeals to Nagel's (1974) explanation of a state's being conscious as involving there being "something it's like" *for* a subject to have the state. It might be held that the "for me-ness" or "subjectivity" that is characteristic of phenomenal states is unintelligible unless the states in some essential way involve a subject (Mcdowell (1997) makes this kind of point). Moreover, surely the "me" for whom different simultaneously unified experiences are "like something", is the same "me" for each experience. It would be very strange to say "this pain is like something for me and this visual experience is like something for me, although I don't mean to suggest that it is the same "me" in each case". How can we make sense of this if we reject the view that they have the same metaphysical subject?

There are two kinds of responses the non-identity theorist might make to this argument. First, she can accept that this "for me-ness" is a serious datum that needs to be explained, but attempt to do it in a way that is consistent with subject non-identity. Second, she could simply refuse to accept that there is a very clearly defined phenomenon here that needs explaining, perhaps giving a deflationary account of the intuitions that may have suggested otherwise.

An example of the first kind of strategy would be an appeal to the self-representational theory of subjectivity, mentioned earlier (e.g. Kriegel (2009)). If every experience involves a self as intentional object, and furthermore it is somehow part of this self-representation that it is the same self in each case, then there is a sense in which there is a single self "standing behind" each experience. But, as explained above, this is perfectly consistent with the subject non-identity thesis, because it is not clear that it need be the metaphysical subject of the experience that is the intentional object of self-awareness (and this is true even if self-representation operates through some kind of token-reflexive rule: the relevant "thinker" of a thought needn't be its metaphysical subject).

The other general strategy involves denying that there really is a well-defined phenomenon here. One view in this ball-park would involve claiming that the closest your experiences get to an intrinsic subjectivity is that they can all automatically be self-ascribed on the basis of introspection. This is not some special

intrinsic feature of the experiences, but rather a consequence of the fact that possession of the first-person concept and the relevant psychological concepts seems to entail an inclination to attribute conscious states to oneself. And as before, this is perfectly consistent with the subject non-identity thesis.

There's a plausible general claim which would support the viability of one of these strategies if it's correct: even if there are strong logical subjects, this couldn't on its own explain the "subjectivity" of experiences. As we have seen, there are plenty of subject-involving events that involve a strong metaphysical subject—for example, your bodily actions at a time may essentially all involve the same metaphysical subject. But these actions do not involve "subjectivity" in anything like the same way that experiences do. So at best, the existence of a strong metaphysical subject would have to be supplemented in some way to explain subjectivity. It's not clear what this supplementation would be; but you might suspect that when it is fleshed out it might reveal that the existence of a strong metaphysical subject isn't doing much of the explanatory work. (This would certainly be the case if the supplementation involved some kind of subject-directed representation, as on a Kriegel-style view).

5.4 Olson's Objection

Olson (1997, chapter 4) compares the view that the there is some body part "directly" involved in the existence of a mental state with the view that there is some body part of a person "directly" involved in their walking, or some part of a machine in a factory "directly" involved in producing an object. He objects to this idea, partly on the grounds that it will be very hard to distinguish in any principled way between parts that are genuinely "directly" involved, and those that aren't.

To offer this as an objection to the Non-identity thesis is to miss the disanalogy between walkings and experiencings that gives us the self-predicament. Being in the self-predicament, we start off knowing almost nothing about the objective features of the self; so to identify the self, it seems that we have no option but to try to isolate those objects that are in some sense "directly" involved in the existence of experience. Even if it is difficult to do this in a principled way, it seems that we must proceed on something like this basis. With a walking, we are already given the metaphysical subject—some person or organism—right from the start, and so to try to whittle down to some smaller object that is "directly" involved in the walking would be to do something bizarre and completely different from what we are doing in attempting to find the experiential self.

5.5 The Objection from Externalism

The view that experiences have brain regions as metaphysical subjects may seem like an extreme form of internalism on which the entire physical basis for an

experience is located within the body, or even within the boundaries of the brain. Is it therefore a view that experiential externalists— or example, naïve realists or phenomenal externalists like Lycan and Tye (Lycan (2001), Tye (2002)) - or fans of "embodied cognition" like Clark, Hurley, or Noe (Clark (1999), Hurley (2002), Noë (2005)) should be lining up to condemn?

Probably not. Even if brain-regions are the metaphysical subjects of experiences, that doesn't mean that these experiences don't involve *relations* between these metaphysical subjects and other items—a point I emphasized above in discussing the subtraction argument. I'm sure typical externalists are thinking of the subject of the experience as the whole organism; but they could be motivated by the Subtraction argument to revise this view. For example, we could have a version of an acquaintance view, on which the acquaintance relation holds between *brain regions* and external objects.

Admittedly, allowing for non-intrinsic experiential properties raises hard questions about how to identify the metaphysical subject of the experience. We can't simply take it to be a fusion of the objects involved in the total realization of the experience (what Chalmers (2000) calls the "total correlate"): if we applied this criterion to a relational view of perceptual experience, we would end up treating the metaphysical subject of awareness as literally having external objects they are perceiving as parts! But if the metaphysical subject can be a mere part of the total correlate, what part? What would be wrong with a view on which the metaphysical subject is a single neuron, or single space-time point, the relevant experiential property being a highly relational one?

These puzzles suggest a kind of deflationary view on which once we have specified the total correlate of the experience there is no substantive further issue about what part of it is the "real" metaphysical subject of the experience. I suspect that the view is too strong, at least in the sense that some ways of describing the experiential state of affairs realized in the total correlate will be more natural or explanatory than others (e.g. one that distinguishes the subject of experience and the target of experience). This issue warrants further discussion.

5.6 The Objection from Functionalism

According to the objection from Functionalism, experiential properties are at least partially individuated in terms of their functional role. When we think about what the functional role of an experiential property is, we will see that different experiential property instantiations within a single conscious life must have the same metaphysical subject (Shoemaker (1985) offers a closely related argument in his review of Parfit's "Reasons and Persons").

There are, I think, two reasons why this might seem like it has to be right. First of all, a standard functionalist theory states the functional roles of different mental properties in a way that presupposes that they belong to the same object.

A typical functional role is given in terms of a theory describing how different combinations of mental states are caused by prior input / mental state combinations, and how they lead to behaviors and other mental state combinations. It is just assumed that these are all properties of a single object. Second, the functional role of a mental property is typically taken to be holistic, in that it is given by its conditional causes and effects *within a whole psychological system.* For it to play this systemic functional role, a whole psychological system must exist. For example, it might be thought that what it is to have a belief that one is experiencing a blue square involves being in a state with certain functional relations to experiences and also to the system of propositional attitudes. Thus, having the belief requires a system of experiences and propositional attitudes to be in place. If the total realization of each kind of mental state is sufficient for the whole psychological system to exist, then the brain areas and other parts of the body involved in each total realization will be the same (or so it might seem). It might be inferred that the metaphysical subjects of different parts of a unified experience must coincide.

In response to this argument, I would accept that Functionalism, as it is traditionally formulated, involves a strong metaphysical subject, but I think we can envisage a modified version of the view that does not require such an assumption. A helpful point to begin with is that in general the functional roles of properties are not given entirely in terms of the conditional causes and effects that the property has on properties *of the same object*—typically, the properties of other objects are involved as well. To use an example from Shoemaker (1984), the property of being knife shaped conditionally causes an object to cut through a stick of butter, the relevant condition being that it is made of steel and is related to the butter in a certain way. This conditional causal power can be part of what defines being knife-shaped, on a functional view of "knife-shaped". Similarly, we might be able to think of the causal transitions within a mental life as involving the mental properties of one object (e.g. a brain area) causing another object to have certain mental properties (e.g. another brain area).

To get a feel for the suggested modification, consider two fairly uncontroversial ways in which we can modify functionalism to avoid a commitment to a strong metaphysical subject. First, consider how a 4-dimensionalist would want to modify the view to allow for the fact that strictly speaking it is different slices of a 4-D worm that are the metaphysical subject of mental states at different stages of a single mental life. She will state the functional theory so that the effects of a total mental state at t include mental states belonging to later time-slices of a single worm, rather than to the same object, strictly speaking. It is hard to see why there is anything problematic about this. Second, consider how a theorist who wants to identify subjects with their brains will state a functional theory that includes the effects of mental states of bodily movements in the theory. In her functional theory, unlike on a more traditional view, the bodily events won't have the same metaphysical subject as the mental events that cause them.

But this is unproblematic because the effects of the mind on the body can be given a functional analysis with the same form as the one we gave for the effects of a knife on a stick of butter.

Bearing these examples in mind, we can see what a functional theory might look like on the Subject non-identity theory. Consider the case of an experience causing a belief to be formed. A standard functional theory might speak of "S perceiving that p causing S to believe that p". On the modified view, we might put it like this "The perceiving part of S perceiving that p causes the believing part of S to believe that p" (note that the believing part of S might just be the whole of S—I am not arguing for subject non-identity for beliefs). Admittedly this sounds very odd, but the correct metaphysics of experience might not correspond very well to our ordinary way of conceiving of transitions within a mental life. It also sounds odd to describe the changing properties of an object as involving the properties of different temporal parts, but that is a pretty weak objection to 4-dimensionalism.

It would be an interesting project to develop further this kind of unselfish functional view. Here is not the place for that—I will rest content having established that functionalism doesn't in any very serious way cut against subject non-identity because there is an apparently viable way to modify a functional theory to accommodate subject non-identity.

6. Further issues for discussion

Having responded to objections, I conclude with some thoughts about how the thesis relates to more general issues about the self and the unity of consciousness.

Perhaps the most important issue for further discussion is the connection between the subject non-identity thesis, and the notion of a "bundle theory" of the self. As I argued above, whether the subject non-identity thesis should be construed as a "no self" view or a "bundle theory" may depend on which self-roles we deem important. For many kinds of "self", the thesis could suggest a "weak bundle theory", on which the "self" depends for its existence on experiences, but experiences do not depend on selves. More needs to be said to justify this claim, including clarifying what versions of the "bundle theory" are conceivable, and how they relate to one another. Here I want to emphasize one important point about what it is to hold a "bundle theory". Even if you thought the self just *is* the stream of consciousness, and you thought that the stream is built up "from below" out of independently existing experiential parts, there would be something potentially misleading about describing it as a "bundle of experiences", because a stream is no mere bundle of experiences—they are unified together in various interesting ways. Moreover, the connotation of "bundling" is that the items in the bundle exist quite independently of one another, their "unity" as bundle being a quite external matter, as if they are tied together by pieces of string. But even if

we accept the subject non-identity thesis, we needn't think of the unity relation as anything like this. In particular, the parts of a unified field of awareness might be highly inter-dependent, even if they have different metaphysical subjects. Two theses that elaborate this idea in different ways are:

> **Unity Internalism**: Phenomenal Unity is an internal relation between experiences.
>
> **The Unity Overlap Principle**: the total realizers of unified experiences must overlap.

Phenomenal unity might be internal in the sense that the facts in virtue of which each of two unified experiences individually exist might be enough alone to ensure that they are themselves unified—no external condition need obtain to link them together as unified. The Unity Overlap Principle is motivated by the thought that what makes different unified experiences conscious might of necessity overlap: the fact that they are unified in part has to do with the fact that there is joint explanation for each being conscious. (I discuss these principles in much more detail in Lee (2014b)).

If such principles hold, there will be an interesting disanalogy between the case of a game of football discussed earlier, and the case of a unified stream of consciousness. The relations between different events in the game in virtue of which they are game-unified are mostly external relations—they have to do with the events being spatio-temporally and causally related to each other in various ways. They are also typically spatio-temporally separated rather than overlapping. The relations that unify experiences together could be much more intimate than this. (Or then again, maybe not—I am not making any claims either way here).

Finally, an issue for further discussion is the relationship between the subject non-identity thesis and attacks on the existence of selves based on appeals to a modular conception of cognition. Dennett (1992) claims that if the brain is really a set of interconnected domain-specific gadgets, then there is no self at the center of a mental world; it might be more accurate to think of the cognition as a collaborative effort between multiple cognitive centers. The Subject Non-identity thesis is only indirectly related to Dennett's version of the No-self idea. First, the Subject Non-identity thesis is only directed at the conscious self, whereas Dennett is taking aim at the subject of *all* cognitive states. Second, Dennett probably need not be read as claiming that "personal-level" mental states belong to different modules, whereas I *am* claiming that something akin to personal-level properties (experiential properties) might properly be thought to belong to different brain areas. Third, Subject Non-identity theorists are not committed to modularity as that is normally understood. They are only committed to the view that the total correlates of experiences do not always completely coincide. Prima facie, someone who held a highly non-modular view of the mind—for example, someone who rejected the existence of *any* spatially or functionally localized, task-specific brain networks—could still accept this.

7. Conclusion

My main conclusions in this paper are as follows: (1) It is important to distinguish the different self-roles that could be used to elucidate what a "self" is, and it is in general a non-trivial matter whether the objects playing different self-roles are the same. (2) We face the "self predicament", which means that it is a hard empirical question what the metaphysical subject of an experience is. (3) Despite (2), there is a good case for the subject non-identity thesis: there are positive reasons for thinking that it is true, and none of the arguments against it we canvassed are particular forceful. (4) Depending on the weight we put on different self-roles, the subject non-identity thesis supports either a kind of "no self" view or a moderate kind of bundle theory; the thesis is therefore of great importance for the traditional debate about the self, and remains important under a wide variety of different conceptions of the "self".[26]

Notes

1. *A Treatise of Human Nature, Book I, Part 4, Section 6, 'Of Personal Identity'* (Hume (2003)). What exactly he meant by this, and specifically whether a "bundle theory" of the self is his considered view, are rightly points of controversy; see Strawson (2011a) for an interpretation of Hume on which he is not a bundle theorist.
2. Mcdowell (1997) and Shoemaker (1985) give forceful statements of this objection.
3. The bundle theory of the self might remind you of the theory that particulars are bundles of universals or tropes, and you might be wondering if consideration of that theory will loom large here. It will not: the "metaphysical subject" of an experience, as discussed here, could be a bundle of universals, or it could be a particular understood in a different way. Also, in my view, the best version of the bundle theory applies only to the basic micro-particulars out of which other particulars are composed—those macro-particulars are not themselves bundles, but mereological composites of micro-bundles. Selves, on any plausible view, are macro-particulars composed of microentities, and it doesn't matter for our purposes what the correct metaphysics of these micro-consituents is.
4. Strawson (2011b) specifically theorizes about what he calls the "minimal subject" of experience defined as "the subject that must remain when everything but experience has been stripped away". One could reasonably interpret him as pointing to the metaphysical subject self-role, as I understand it here, although I don't define it in this way (see section 1 below).
5. Some theorists will think that the plural event is identical with such a singular event. Such an identification is probably, in my view, less problematic than identifying a material thing with its parts (see McDaniel (2008) for an interesting argument against this view). However, whether or not this is correct won't matter in what follows: the events are clearly equivalent at least in some weaker sense.
6. Of course such a theory will need to account for the relations between experiences in virtue of which they are unified into a single stream. This is no objection to

the view, however, since the subject-centered view will need such an account also. (The subject-centered theorist could try the view that simply belonging to the same metaphysical subject explains unity, but this is probably a non-starter). I discuss this issue in detail in Lee (2014b).

7. This is one respect in which the view can be seen as a kind of "bundle theory". Although just like other experiences, a total experience can *also* be treated as having a single metaphysical subject, as discussed above. Compare a plural event like a game of football. We could define an object SUM that is a mereological sum of all the objects taking part in the game: the players, the stadium, etc. We could then treat all the events in the game as if they have SUM as their metaphysical subject; for example, instead of talking about player X scoring a goal, we could talk about the event of SUM having X as a part, and X scoring a goal. Clearly the normal description of the game as a plural event is more fundamental here. The subject non-identity theorist thinks something is true of a total experience, even though we might normally think of it as having a single logical subject.
8. A possible exception to this: the temporal properties of the subject.
9. We need to be careful not to overstate this point. Arguably, our introspective awareness of experience does reveal to us "topic neutral" facts about experiences, such as facts about their causal impact on behaviors like verbal report, facts that will be critical in figuring out what the neural correlates of the experiences are.
10. Perhaps we don't need to know exactly which physical properties are correlated with experiential properties, but just enough about these properties to objectively identify which object has them. For example, perhaps having knowledge of the causal role of a relevant physical property would be enough. What I say here probably needs to be qualified to allow for this.
11. This might even be viable on a property dualist view, where we are interested in which object primitively instantiates some phenomenal feature. Even in this case, the most elegant theory of the nomological connections between the physical and phenomenal realms might pick out certain particular objects as the metaphysical subjects of the relevant experiences.
12. More generally, I suspect that we can't decide what the correct analysis of experience relativization is without first deciding whether body parts can be said to have experiential properties in a more fundamental sense that whole organisms.
13. The one reasonable worry one might have about (1) concerns *uniqueness*—maybe the existence of an experience is overdetermined in such a way that there are multiple minimal sets of events that would be sufficient on their own for an experience of the same type as E to occur. This is more of a technical problem than a serious philosophical objection, however. For example, we could get round it by taking SUB to be the *sum* of these multiple minimal sets.
14. One could object to (2) that the detached body part would be identical with my whole body in the counterfactual scenario, not a proper part of it—that is, my body can be variably constituted by different matter, including matter that is currently constituting only a part of it. On this view, (2) is false, and (3) is true (thanks to Nick Stang for this objection). In response, one could reformulate (2) and (3) in terms of the matter or material parts that constitute the metaphysical subjects of E and C(E). (2) will say that the subject of C(E) is constituted by matter that constitutes a proper part of S's body, and (3) will say that C(E) and

(E) have subjects constituted by the same matter. The defense of (3) that follows can easily be adapted to this version (see also footnote 15).

15. Some theorists (such as the objector in footnote 15) may want to distinguish a chair from the hunk of matter that the chair is made from. They might put the points here somewhat differently: they might say that it is a matter of whether the property of *constituting a chair* is an intrinsic property of a hunk of matter, not whether *being a chair* is an intrinsic property of a chair. Everything I say here about intrinsic properties could be restated in these terms. Also, if we define an intrinsic property as one whose instantiation supervenes on the perfectly natural properties of, and relations between, the parts of the object (see below, and footnote 16), then being a chair *will* come out as an extrinsic property, even on a view that distinguishes the chair and the chair-shaped matter.
16. The apparent circularity here can be avoided by appealing to fundamental, or "perfectly natural" properties and relations (Lewis (1983)). An intrinsic property is one that is grounded in the perfectly monadic properties of, and perfectly relations between, the parts of the object.
17. I stipulate that an experience is *associated* with an object O, just if O has natural spatial boundaries, and the metaphysical subject of the experience exists within the spatial boundaries of O.
18. Although admittedly, on some versions, the relevant functional role might involve the organism's body, and therefore seem to require the experiencer have a body; however, proponents of functionalism have typically taken this to be an objection, and tried to formulate their view so that experiences aren't body-requiring (e.g Shoemaker (1976)). Furthermore, functionalist theories can be formulated in a way that doesn't require that the metaphysical subject of experience be coincident with a human body, even when the brain *is* connected to one—see below, 5.6.
19. Another caveat—some experiential states may seem to require that the metaphysical subject of experience have a body with natural boundaries, because they are *directed* towards that body, as in cases of proprioceptive awareness. For example, perhaps proprioceptive awareness involves a subject standing in an acquaintance relation to their body, and so requires the body to exist. The point to make here is that this does not really suggest that such properties are boundary-constraining, because awareness of one's body might be just like awareness of any other material object. Just as awareness of a chair doesn't suggest that the metaphysical subject of experience has chair-shaped natural boundaries (even if it implies the existence of some *other* object that does), awareness of one's body doesn't suggest that the metaphysical subject of the awareness is coincident with the body, rather than, say, a part of the body like the brain.
20. In defense of Total Realization Holism, one important point to note here is that even if it is true, there can still be a sense in which a visual experience that is unified with an auditory experience is more closely associated with the visual system than the auditory system. The total realization holist can appeal to Shoemaker's notion of core realization to elaborate this point. The core realizer of an experience is a set of events that differentiate an experience of a certain kind from other actual or possible experiences; for example, even if a visual experience isn't realized totally in the visual system, it might still be true that the events that determine what kind of visual experience it is are totally contained within the visual system. A total realization holist could say that a visual and auditory

experience have different core realizers, even if they have the same total realizer. (They could even gloss this by distinguishing between the "core self" and the "total self" associated with an experience, and hold that core selves within a stream are, in general, quite variable).

The possibility of this move suggests that the total realization holist can avoid the highly implausible view that there is nothing about the realization of a visual experience accompanied by an auditory experience that associates it particularly with the visual system. Nonetheless, it remains true that total realization holism is implausible. Not only is an auditory experience not core-realized in the visual system; it also seems plausible that its total realization doesn't involve visual events either, even if they are involved in the simultaneous realization of a visual experience.

21. You might think that even if "being conscious" is holistic in this way, that won't imply total realization holism. Wouldn't the total realization of, say, a visual state involve the existence of a certain visual mental representation, and the representation having a certain existential or general relational property, such as the relational property of there being no other representations in the system with a stronger activation level? This would be different from the total realization of an auditory experience, which would involve an auditory representation with a different existential relational property.

    The objection raises the issue of how exactly we should understand "realization" here. I think in this context it makes sense to understand realizers as specific rather than general, which would rule this kind of thing out from being a realizer. That is, the relevant realization would involve the specific facts that make the existential or general fact true. For example, if the existential fact is the fact that there are no other representations that are activated more strongly, then the specific facts would involve the facts about the other representations in virtue of which this is true.

    I do not say that there aren't important questions about how to develop this notion of a "specific realizer" of an experience, but I think it would be uncharitable to the realization holist to press this objection to them. It is analogous to the objection that the realizer of an object's being between 10 and 15 kg isn't the realizer of its being 12kg, on the grounds that a certain abstract existential fact about the object might be sufficient for it having the determinable property, but not the determinate property. Even if there is a sense of "realization" on which this is true, it obviously doesn't show that the objects having the determinable property doesn't depend in an important sense on its having the determinate property.

22. There is a general problem with this idea worth mentioning. Even though "I" is governed by the apparently straightforward token-reflexive rule (it refers to whoever or whatever produces a token of "I"), it is often extremely unclear to what it refers, because there are many candidates for being the relevant speaker or thinker: a human animal, a Lockean person, a body, a brain part, etc. The token reflexive rule on its own may not point us to the referent of "I" without some supplementation, either in the form of a *different* self-conception that is implicit or explicit in our use of "I", or perhaps some metaphysical view that restricts the candidate "I" thinkers. It is very hard to provide such supplementation in an uncontroversial way; indeed, one reasonable response to the problem would be to

hold that it is typically highly indeterminate to what "I" refers. For these reasons, we might think that other self-roles have a more central role to play in the debate about the self. (See Johnston (2011) and Sider (2001b) for related discussion).

23. Note the interesting possibility of an error theory here: even if there is ubiquitous self-awareness, it is possible that this self is a merely intentional object, i.e. that there is nothing in the real world towards which our self-awareness is directed.
24. Two qualifications to this should be mentioned here. First, if the relevant "self" isn't identical with the metaphysical subject of an experience, but is a *proper part* of such a metaphysical subject, then the experience *will* depend for its existence on this "self". But I take it that this probably isn't true for any of the "selves" I considered here, and so is not relevant. Second, and more relevantly, one way in which a mental state might be individuated in terms of some object, other than having it as metaphysical subject, is by having it as an intentional object. A mental state that involves self-reference or self-awareness might thereby involve the "self" in some sense. The subject non-identity thesis doesn't rule out some mental states being self-individuated in this sense. This raises the issue of whether the "self" that these states are directed towards could exist partly in virtue of these self-directed mental states existing. A kind of non-reductive bundle theory might hold that the selves and self-directed states are mutually inter-dependent. But one could also maintain that the self-directed states *are* more basic than the selves, despite being self-directed; for example, one could hold that there is a way of understanding their "selfish" content that doesn't require specifying the self towards which they are directed. This issue certainly deserves more discussion. (Parfit (1999) appears to give up his earlier, more reductive view of the self based on similar considerations).
25. If the subject non-identity thesis generates a concern for Descartes, it is surely this: that if we allow that you might not be the metaphysical subject of experience but rather some larger object that merely has the experience going on inside it, then the existence of your experience seems to *fail to guarantee that you exist*. By contrast, the thesis doesn't seem to generate a special new worry about how you know that these experiences are yours and not someone else's.
26. Thanks to David Chalmers, Matti Eklund, Rory Madden, Nick Stang, and participants at the Arizona metaphysics workshop, and the Paros personal identity workshop for helpful comments and discussion.

References

Bayne, T. (2010). *The Unity of Consciousness*. Oxford: OUP.

Bayne, T., & Chalmers, D. (2003). What is the unity of consciousness? In Cleeremans, A. (ed.) *The unity of consciousness: Binding, integration, and dissociation*, 23–58. Oxford: OUP.

Block, N. (2003). Mental paint. *Reflections and replies: Essays on the philosophy of Tyler Burge*, 165–200. MIT Press.

Campbell, J. (2002). *Reference and consciousness*. Oxford: Clarendon Press.

Cassam, Q. (1999). *Self and World*. Oxford: OUP.

Caston, V. (2002). Aristotle on Consciousness. *Mind*, 111(444).

Chalmers, D. J. (1998). On the Search for the Neural Correlate of Consciousness. In Hameroff, S. R., Kaszniak, A. W., & Scott, A. (eds.), *Toward a science of consciousness II: The second Tucson discussions and debates*. p. 219. MIT Press.

Chalmers, D. J. (2000). What is a Neural Correlate of Consciousness? in T. Metzinger, (ed.) *Neural Correlates of Consciousness: Empirical and Conceptual Questions*. MIT Press.
Clark, A. (1999). An embodied cognitive science? *Trends in cognitive sciences*, 3(9), 345–351.
Dennett, D. (1992). The Self as the Center of Narrative Gravity. In Kessel, F. S., Cole, P. M., Johnson, D. L., & Hakel, M. D. (eds.) (1992). *Self and consciousness: multiple perspectives* (Vol. 6). Lawrence Erlbaum.
Dretske, F. I. (1997). *Naturalizing the mind*. Bradford Books.
Hurley, S. L. (2002). *Consciousness in action*. Harvard University Press.
Hume, D. (2003). *A treatise of human nature*. Courier Dover Publications.
Jackson, F. (1977). *Perception: a Representative Theory*. Cambridge, England: Cambridge University Press.
Johnston, M. (2011). *Surviving death*. Princeton University Press.
Kim J. (1976). Events as Property Exemplifications. In M. Brand and D. Walton (eds.), *Action Theory*, Dordrecht: Reidel, pp. 159–77.
Kriegel, U. (2009). *Subjective Consciousness, a Self-Representational Theory*. OUP.
Lee, G. (2014a). Temporal Experience and the Temporal Structure of Experience. Philosopher's Imprint, 14(3), pp. 1–21.
Lee, G. (2014b). Experiences and their parts. In Bennett and Hill (eds.) *Sensory Integration and the Unity of Consciousness*. MIT Press.
Lewis, D. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, 61(4).
Lichtenberg, G. (1971). *Schriften und Briefe*, Vol. ii. Munich: Carl Hanser Verlag.
Lofting, H. (1998). *The Story of Doctor Doolittle, being the History of his Peculiar Life at Home and Astonishing Adventures in Foreign Parts*. London: Macmillan.
Lycan, W. G. (1996). Consciousness and Experience. Cambridge MA: Bradford Books.
Lycan, W.G. (2001). The Case for Phenomenal Externalism. *Noûs*, 35(s15), 17–35.
Martin, M.G.F. (2003). The Limits of Self-Awareness. *Philosophical Studies*, 120(1), 37–89.
McDaniel, K. (2008). Against Composition as Identity. *Analysis* 68: 128–33.
Mcdowell, J. (1997). Reductionism and the First Person. In *Reading Parfit*, ed. Jonathan Dancy. Oxford: Blackwell.
Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, 83(4), 435–450.
Noë, A. (2005). *Action in perception*. MIT Press.
Olson, E. (1998). There is no problem of the self. *Journal of Consciousness Studies* 5, 1998, 645–57.
Olson, E. (1997). *The Human Animal*. New York: Oxford University Press.
Papineau, D. (forthcoming). Sensory Experience and Representational Properties. *Proceedings of the Aristotelian Society*, 114 (1).
Parfit, D. (1999). Experiences, subjects, and conceptual schemes. *Philosophical topics*, 26(1/2), 217–270.
Peacocke, C. (1995). Demonstrative Content: A Reply to John McDowell. *Mind*, 1995: 126.
Prinz, J. (2012). *The Conscious Brain*. OUP.
Rosenthal, D. M. (1990). *A theory of consciousness*. Zentrum für interdisziplinäre Forschung.
Russell, B. (1912). The Problems of Philosophy. New York: Henry Holt and Company.
Schaffer, J. (2010). Monism: The Priority of the Whole. *The Philosophical Review* 119 (1):31-76.
Shoemaker, S. (1976). Embodiment and Behavior. In A. Rorty (ed.), *The Identities of Persons*. Berkeley University Press.
Shoemaker, S. (1984). Causality and Properties. In *Identity, Cause and Mind*. Oxford: Oxford University Press.
Shoemaker, S. (1985). Critical Notice of "Reasons and Persons". *Mind* 94 (375):443-453.
Shoemaker, S. (2007). *Physical Realization*. Oxford: Clarendon Press.

Sider, T. (2001a). Maximality and Intrinsic Properties. *Philosophy and Phenomenological Research* 63.
Sider, T. (2001b). Criteria of Personal Identity and the Limits of Conceptual Analysis. Philosophy and Phenomenological Research.
Strawson, G. (1999). The Self and the SESMET. Journal of Consciousness Studies.
Strawson, G. (2011a). *The Evident Connexion: Hume on Personal Identity*. OUP.
Strawson, G. (2011b). The Minimal Subject. In *The Oxford Handbook of the Self*, ed. Gallagher, S. OUP. pp. 253–278.
Strawson, P. F. (1966). *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason*. Routledge.
Tye, M. (2002). Representationalism and the Transparency of Experience. *Nous*, 36(1), 137–151.
Tye (2003). *Consciousness and Persons: Unity and Identity*. MIT Press.