3

Subjects and Consciousness

Christopher Peacocke

What is it to represent oneself as oneself? And what is the nature of the entity, if any, so represented? These are the two questions to be addressed in this chapter. The first is a question about the nature of intentional content; the second concerns the metaphysics of the self. We should aim for an integrated pair of answers to these questions, an integration that makes it clear why the content elucidated in answering the first question is a way of representing an entity of the sort characterized in answering the second question. In this particular area, I believe that reflection on the question about content yields some important resources for answering the second question. So I will begin with the question about content.

I

A creature may see something as coming towards it. It may have an apparent memory of being involved in an earlier encounter at a particular location. It may have an apparent action-awareness of moving its head. All these conscious events have a content concerning a subject. The content in question has an intermediate status. The content is more than merely one which concerns the object that is in fact the subject. The content involves less than the full conceptual first person content. Merely seeing something that is in fact the subject includes the case of seeing something that is in fact oneself, say in a mirror, without one's realizing that it is oneself. That can occur without the object being in any way represented as oneself in the perceptual experience. Seeing something as coming towards one involves much more than merely reference to something that is in fact oneself. Yet seeing something as coming towards one is something that can occur in subjects who lack concepts. Concepts are constituents of the intentional contents of states that subject can enter for reasons; concepts are constituents of the contents of judgements; it is in the nature of concepts that they are constituents of the states and events in which a subject displays a sensitivity to reasons. It seems that first person perceptual events, memories, and action-awarenesses of the sort we have just mentioned can be present in creatures that have only a more

primitive system of nonconceptual representations of the world, and who do not operate at the level of reasons at all. So the question arises: what is the nature of the self-representation in the content of these conscious events?

It is in the nature of the type of content of each of the events of this sort that their correctness conditions concern the subject of the event. A particular token perceptual experience of an object as coming towards one has a correct content only if the object in question really is coming towards the subject of the experience. A subject who has an apparent memory of an encounter with a certain kind of animal at a certain location has a correct memory only if the subject enjoying the apparent memory really did have such an encounter. A subject's apparent action-awareness of moving his head is correct only if the subject enjoying the apparent awareness really is moving his head. As is well known, in all three cases, causal materials must be added if these necessary conditions are to be strengthened to reach the status of genuine perception, memory, and action-awareness.

When I say that "it is in the nature of the type of content" that the correctness condition concerns the subject of the events, I mean at least the following: that no further information about the reference of the ways things, events, and properties and relations are given in the intentional content of these events is needed to settle whether or not the content refers to the subject of the event. The way in which the subject is given in the awareness entirely settles that it is the subject of the awareness that the content concerns. In this it differs from any perception that presents a person, and which happens to refer to the subject who enjoys the perception. The fundamental reference rule for any instance of the perceptual-demonstrative type *that F* (such as *that man*), where the demonstrative is tied to an experience in which something is apparently presented in a given way W, is this: the demonstrative refers to the man that is perceived in way W. This fundamental reference condition implies that the reference of the perceptual demonstrative is not necessarily the subject who is enjoying the perceptual experience in question.

Actually we need a stronger formulation of the point. In the contrasting conceptual case, we can conceive of a mixed descriptive-demonstrative content such as *the agent of this thinking*, or *the owner of this experience*. There is clearly a sense in which it is in the nature of the type of these contents too that they refer to the agent of the thinking and to the owner of the experience respectively. At the conceptual level, neither of these is equivalent to the first person concept. Although the referents of these complex descriptive-demonstrative concepts are guaranteed to be, respectively, the subjects of the thinking and of the experience, the determination of this reference goes via a component *the agent of* and *the owner of* that features in the content itself. For the genuine first person, by contrast, the determination of the reference as the subject does not go via such a descriptive component in the content itself.

We must respect the distinction between complex descriptive-demonstrative notions and the genuine first person if we are to characterize properly such states as the phenomenology of ownership. In his paper "Self-Consciousness and the Unity of Consciousness", Tim Bayne writes, "I take the sense of ownership to be an experience

whose representational content is roughly: this (target) experience is had by the subject of this (reflexive) experience" (Bayne 2004, p. 231).[1] It seems to me that this understates the sense of ownership, which involves the sense that the experience in question is *mine*, and not merely something had by whoever may be the subject of this experience. We cannot properly capture the sense of ownership without mentioning the first person notion as part of the (nonconceptual) content of the subject's consciousness. Any syndrome in which someone experiences a pain but does not experience it as meeting the condition *This is mine*, is a syndrome in which the subject lacks a sense of ownership of that particular pain, whatever else she may represent about whoever is the subject of the experience.

So in the requirement that it be in "in the nature of the type of content" of the event or state in question that it refer to the subject of the state or event possessing the content, the following should be understood: that the determination of the subject as the reference does not proceed via satisfaction of some other condition that is involved in the nonconceptual content. It may help to point out that the distinction being invoked here is an analogue for these contents of Saul Kripke's distinction at the linguistic level between *de jure* and de facto rigid designation.[2] The first person nonconceptual content of a mental event is of a type whose instances refer *de jure* to the subject of the mental event in question.

These points are at the level of type of content, rather than the contents themselves. The fact that they are at the level of the type permits us to sidestep an objection. We can imagine an objector who believes in *de re* senses, and in their analogues at the nonconceptual level.[3] This objector might take a case like the one mentioned by Ernst Mach in which he sees a reflection—in fact of himself—in the window of a bus, and represents the object using the perceptual-demonstrative "that shabby pedagogue" (Mach 1914, p. 4). The believer in *de re* contents might develop his position in such a way that he insists both that: this particular *de re* content "that shabby pedagogue", as employed in thought on this occasion, has Ernst Mach as its reference essentially; and that only Ernst Mach could be the subject who enjoys the particular token perceptual visual experience that makes the perceptual demonstrative available. We need not for present purposes enter discussion of the former (contentious) claim, because at the level of types, nothing analogous can be said of the perceptual-demonstrative type *that shabby pedagogue*. That is, it is not in general true that any intentional content of the perceptual demonstrative *type* "that shabby pedagogue" will also refer to the subject who enjoys the relevant perception of the shabby pedagogue. By contrast, it is in the nature of the first person type, as it occurs in the intentional content of some mental event, that the

[1] Bayne could certainly change his formulation in the way I will be recommending consistently with (and arguably strengthening) the rest of the arguments in his paper, whose main points seem to me very well taken.

[2] See his introduction to the book version of *Naming and Necessity* (1980).

[3] On *de re* or "object-involving" senses, see G. Evans (1982); J. McDowell (1984); and on the relation between types and instances of types in the realm of the senses, see C. Peacocke (1981).

instance so occurring refers to the subject of that same mental event. The points I am making about the first person type, whether conceptual or nonconceptual, are then orthogonal to the issue of whether or not there are *de re* contents.

De jure truths about the nonconceptual first person must have a source or explanation. What is it? One natural answer to this question is this:

> what makes a component of nonconceptual content something of the first person type is that the fundamental condition for an instance of the type to refer to something, when it is in the content of mental state or event M, is simply that it refers to the subject of M (M's owner).

We can call this *the subject-constitutive hypothesis* about the nature of the nonconceptual first person.

The subject-constitutive hypothesis does not imply that the nonconceptual first person is some disguised complex descriptive-demonstrative content. We must always distinguish the material involved in the reference rule for a concept (or nonconceptual content) from the concept (or content) itself. The conceptual content *now* is individuated by the rule that in any thinking, it refers to the time at which that thinking occurs. But concept *now* itself is not structured. It should not be identified with the complex descriptive-demonstrative *the time of this thinking*. Nor should the nonconceptual first person—*i* as I will label it—be regarded as identical with some complex content *the subject of this state*. The nonconceptual content *i* itself is unstructured.

There is a consequence of the fact that an instance of *i* refers *de jure* to the subject of any state in whose content it features. Suppose a particular conscious subject *s* is in a mental state with the nonconceptual first person content. (My apologies for the use of the variable, but we need it to make the point sharply.) Then our subject *s* is in a mental state that *de jure* represents something about *s*. Our subject is a self-representer. Our subject has the property

λy [y represents something about y];

and our subject has this property in virtue of the nature of the type of content of his mental state. (Here I follow standard notation: λx[Fx] is the property of being F; λx λy [Rxy] is the relation R; and so forth).

In his important discussion of Elizabeth Anscombe's view that "I" does not refer at all, Gareth Evans notes that some of her points can be met by observing that the word "I" is a device that each person knowingly and intentionally uses to refer to himself. That is, each person knowingly and intentionally has the property

λz [In using "I", z refers to z],

the property of being a self-referrer.[4] This distinguishes "I" from any proper name, or, I would say, any other expression or concept other than the first person. Now we have

[4] See the appendix to ch. 7 of G. Evans (1982).

just observed in the preceding paragraph that our subject in a nonconceptual first person state has the property λy [y represents something about y], and does so as a result of the very nature of the content. This shows that a form of Evans's point applies equally at the nonconceptual level, at a level below that at which we can properly speak of the intentional and knowing use of either words or concepts.

Whatever may be the explanation of the *de jure* truths about *i*, we can say that the (nonconceptual) content of the events and states in the intermediate cases in which we are interested is *intrinsically subject-referring*. It will be very convenient to have a label for this phenomenon. Suppose a particular subject *s* and one of its mental events *e* stand in the following relation, and do so as a result of the nature of the type of content *e* in the way we have been discussing: λx λy [x is the subject of y & the content of y refers to x].

Then when that condition is met, I will say *e* stands in the relation of *subject-reflexivity* to the subject *s*. The same applies to states as well as to events. More briefly, I will often speak of the mental event *e having the property of subject-reflexivity* if there is some subject to which *e* stands in the relation of subject-reflexivity.

In the most recently displayed formula, within the square brackets, the variable 'x' occurs twice in the characterization of subject-reflexivity. That is an initial identification of a respect in which these contents involve a kind of subject-reference in the subject's mental event. It is at the very least an open question whether or not the subject of any such event must thereby be employing a first person concept of the subject. I will be arguing that the subject can enjoy such an event without possessing the first person concept. The property of being subject-reflexive in the sense characterized is a generic notion of an event or a state's having a *de se* content. It can apply both when an event or state's content is conceptual, and when it is nonconceptual.

A subject-reflexive state or event can have a content that refers to the subject of that state or event without the subject also being given at the same time in some other, further way—be it perceptually, or in some other demonstrative fashion made available by some conscious state. To be a subject-reflexive state or event, it suffices that the state or event's content be of a type whose instances refer *de jure* to its subject. Whether a subject enjoying mental states with *de se* contents also has to have other background capacities or representations of the world, or must conceive of the world as being of a certain kind, are questions to which I will return.

It is the subject-reflexivity of the state involved that we need to highlight if we are to characterize adequately what is distinctively required by the English locutions "remembering being F", "remembering doing so-and-so", as distinct from "remembering my being F", "remembering my doing so-and-so". As James Higginbotham and Michael Martin have remarked to me, the latter pair of locutions can apply even when remembered from the third-person point of view—for instance, you see oneself in the mirror or on closed circuit TV in the memory. The former pair, "remembering being F" and "remembering doing so-and-so", can be applied correctly only when the subject has a memory with the property of subject-reflexivity. You may remember your conducting the orchestra if you remember seeing the then-live video feed of your

doing the conducting. But you remember conducting the orchestra only if you remember the conducting from the point of view of the conductor, with the orchestra in front of you. The utility of the linguistic form "remembering ϕ-ing" is precisely that it allows us to pick out exclusively memories that are subject-reflexive in respect of the ϕ-ing, in a way the locution "remembering my ϕ-ing" does not.[5]

We could give an entirely parallel explication of the similar intermediate case of a present tense but nonconceptual analogue of *now*, one that features in the contents of apparent perceptions and action-awarenesses. For events with such an intermediate kind of present-tense content, the following relation holds between the time t such an event e occurs, and the event e itself, and the relation holds as a result of the nature of the type of content, without reliance on further information about the case:

$\lambda x \lambda y$ [x is the time of occurrence of y & the content of y refers to x].

This can equally be part of an initial identification of a respect in which these events can have a distinctively present-tense content without apparently requiring the subject to have a conceptual constituent *now*. We could similarly call the relation and property in question in this temporal case that of *time-reflexivity*.

In central cases, when a subject has an experience of, say, having a pond to his left and also has an action-awareness of running straight ahead, the subject represents himself as instantiating the conjunction of these two properties. This raises a question. Under the characterization I have given, the fact that the subject is in a position to represent himself as having the conjunctive property does not immediately follow from his being in a subject-reflexive state that he has a pond to his left and is in a subject-reflexive state that he is running straight ahead. From the fact that someone is in a subject-reflexive state that concerns object x, and is at the same time in a subject-reflexive state that concerns y, where in fact $x = y$, it by no means follows that the subject is in a position to appreciate that it is one and the same thing that is in both states. Though a creature can be in a subject-reflexive state that represents it as F and also be in a subject-reflexive state that represents it as G, nothing in what I have said so far has explained how the subject is in a position to register that he is in a subject-reflexive state of being both F and G.

Now subjects must in fact be capable of integrating the contents of those of their conscious states that exhibit subject-reflexivity into such conjunctive representations. For a subject who possesses and exercises the first-person concept, it is unproblematic how this could be done. A perceptual experience which represents the subject as having a pond to his left entitles the subject, other things equal, to judge a conceptual content of the form *that pond is to the left of me*. This content contains the first-person concept, with 'me' as the accusative form of the English expression of the first person

[5] The false claim that "I remember being at the meeting" is equivalent to "I remember that I was at the meeting" is implied in A. Prior (2003, p. 225).

concept. The judged content *that pond is to the left of me* is then suitable for inferential integration with other first-person contents, such as *I am running quickly straight ahead*. A thinker who accepts these two conceptual contents will, as a result of an inference of conjunction-introduction, be in a position to self-ascribe the property of both having a pond to the left of him, and of running quickly straight ahead. This solution is evidently not available for subjects who do not have the first-person concept, yet who nevertheless succeed in integrating representations about themselves. It is quite implausible that the integration of the contents of such representational states is restricted to creatures who possess concepts (if we admit a conceptual/nonconceptual distinction at all). So we need a different explanation of this phenomenon.

When a subject has a perceptual experience of being *F* and, say, an action-awareness of being *G*, normally representations of those two properties each enter an object file on the subject. An object file is a store of mental representations whose contents are all taken, in one way or another, to apply to the same thing. The idea of an object-file has been used in the explanation of propositional-attitude phenomena by a series of writers including Paul Grice, Michael Lockwood, Peter Strawson, and Robin Jeshion (cf. Grice 1969; Lockwood 1971; Strawson 1974; Jeshion 2002). It has also been used in the explanation of perceptual phenomena by Daniel Kahneman (cf. Kahneman et al. 1992). In the propositional-attitude case, the taking to apply to the same thing is at the level of belief and judgement. In the perceptual case, it is a matter of the perceptual system representing to the subject one and the same thing as having several current properties, including relational properties. Some apparatus takes information from the subject's various sensory, perceptual, and action systems, and integrates that information by placing predicative materials drawn from various sources into the subject's file on itself. That will determine the subject's present-tense awareness, his experience of how things are with him now.

There is, however, a fundamental difference between the operation of a subject's file on itself and perceptual object files, even if we consider only the nonconceptual level. To articulate the difference, I consider first perceptual object files.

A file on a perceived object is indexed by where, egocentrically, the object is perceptually given. It contains information about the object's currently perceived properties. Human perceptual systems also have the ability to keep track of where the object has been recently, the location from which it has travelled to its current position. Humans can do this for several objects simultaneously, as one of Zenon Pylyshyn's well-known demonstrations shows.[6] So as time passes from t_1 to t_2, in the case of perceptual object files, the system has to accomplish two tasks. One task is to form new object files, with information about the current properties of the object at a given location is at the new time t_2. The other task is to achieve representations of identities over time between the objects currently perceived at this later time t_2, and the objects as perceived as being at particular places at the earlier time t_1. Which of these identities hold is an entirely

[6] See the demonstration online at http://ruccs.rutgers.edu/finstlab/MOT-movies/MOT-Occ-baseline.mov (accessed 1 August 2010).

empirical and contingent matter. The object that is at a particular location at the later time might earlier have been at any one of many other locations.

By contrast, as time passes in the first person case, nothing of quite the same kind is required as is needed for the second task in the perceptual case. If a subject at t_1 has a nonconceptual representation of itself as *f*, by means of a file on itself, it suffices to update this at t_2 to a representation that at the earlier time, it was *f*. (I continue to use lower case italics for nonconceptual contents.) This past tense predicate can be combined with other present tense predicates in the subject's file on itself to yield representations to the effect that the subject was *f* and is *g*. In contrast to the perceptual case, here there is no empirical, contingent identity over time that needs further determination. If the earlier representation of itself as *f* was correct, then when the later representation is updated by the appropriate change of tense, so will the representation updated at t_2 also be correct. This holds unconditionally, and not merely in normal circumstances. It holds unconditionally because the later representation is a representation by, and about, the same subject as the earlier representation is about.

This contrasts sharply with the effect of similarly pure temporal updating of the object file labelled "the thing at egocentrically identified location *p* at t_1". Suppose a predicate with the meaning *f* is in the perceptual object file so labelled. (Actually it would be labelled using a form of the present tense.) Simply changing, as time passes from t_1 to t_2, the temporal parameter from *now* at t_1 to *a moment ago* at t_2 in no way ensures the preservation of truth-value. It may be a different object at the location *p* at the later time t_2 than was there at the earlier time t_1. That different object may not at t_2 have been *f* a moment ago. That is why a tracking mechanism is also needed in the perceptual case. It performs the second of the two tasks we mentioned.

In the case of perceptual object files, Pylyshyn has a theory of what he calls 'FINSTS', 'fingers of instantiation', subpersonal pointers that keep track of a perceived object over time.[7] When the same subpersonal finger of instantiation points to an object over time, it is experienced as the same object over time. On Pylyshyn's account, the explanation of our ability to keep track of an object over time involves these FINSTS. Perceptual object files must be supplemented with FINSTS if we are to explain our abilities to keep track of a particular object over time. My point is that in the first person case, there is no need for an analogue of FINSTS. Pure temporal updating of the subject's file on himself suffices in a way in which it cannot suffice in the perceptual case.

The argument developed in the preceding three paragraphs is essentially an adaptation, both to the nonconceptual case and to the level of mental representations, of points made about the updating of conceptual first person beliefs by Evans. In *The Varieties of Reference*, Evans discussed the relation between a later disposition to judge "I was previously *F*" and its relation to a present tense judgement "I am *F*" made at an

[7] For an overview of much previous work, see Z. Pylyshyn (2007, ch. 1).

earlier time. He wrote, "as far as the 'I'-idea is concerned, the later dispositions to judge flow out of the earlier dispositions to judge, without the need for any *skill* or *care* (not to lose track of something) on the part of the subject" (Evans 1982, p. 237). I suggest that the explanation of the 'flowing out of the earlier dispositions' that Evans mentions here is the phenomenon of pure temporal updating of the subject file, together with the constitutive links between the nonconceptual first person and conceptual " 'I'-idea".

Whenever an animal or a human has the ability to distinguish in one way or another between the case in which it is the same and the case in which it is a different object which is now *F* and was *G*, there must be an explanation of how the animal or human has this ability. In the case of the perceptual distinction, Pylyshyn's FINSTS provide us with a possible and plausible explanation. In the case of distinguishing from the first person point of view whether it was oneself who was recently *F*, pure temporal updating of the subject file, as part of the mechanism that places past tense contents in the subject file, explains the capacity.

In each of these two very different kinds of case, the underlying explanation of the ability in question involves a sensitivity to identity without requiring at the explaining level a further representation of identity. We do not have to check, at the conscious personal level, whether or not it's the same FINST as earlier that's doing the pointing. That way infinite regress lies: for how would the system determine sameness of FINST? That itself is another identity question, and it is a mistake to think that it is a question the system must somehow address. These questions must come to an end at some point if they are ever to be answered by the system, and it is sensitivity without further representation that makes this possible.

The same applies to the operation of entry mechanisms for past tense predicates into the subject's file on himself. There is no question of the subject, at the personal and conscious level, checking whether the pure temporal updating is being done correctly, or is being done on the file with the right labelling. Again, that way regress lies, a regress that is entirely avoidable. And once again, a mechanism that is sensitive to identity does not need to have some independent test or criterion for identity.

None of these mechanisms is infallible. They may all fail or misfire in various circumstances, in which cases they respectively fail to represent, or misrepresent.

In an effort to make clear the central difference between the updating of the subject file on the one hand and perceptual object files on the other, I have used the simple notation '*f*' for the predicative component of a nonconceptual content. The simplicity of the notation should not mislead. In a wide range of cases, the representations that enter the subject's file on itself will be those appropriate to the distinctive rich spatial contents of perception and action-awareness. Just as these states and events have spatial content (that I tried in earlier writing to capture by the notion of scenario content), so do autobiographical memories of these very states. So a subject's file on itself should not be thought of as some subpersonal analogue merely of a set of predicates, some with past-tense parameters. It should be thought of as including a rich array of imagistic representations. They will be prime candidates for representations that enjoy what

Roger Shepard and Susan Chipman describe as standing in second-order isomorphisms with the reality they represent (cf. Shepard and Chipman 1970).

We can distinguish various categories of object file by what explains how a representation enters an object file of that category. For belief files associated with names (in a given connection), what explains a sentence entering that file is the thinker's acceptance of some sentence containing the name (in that connection). The sentence might be "Paderewski played Chopin rather fast". Such acceptance is something potentially under the subject's rational control, as the agent of his thinking; it is at the personal level; and it can be conscious.

By contrast, the conditions for entry of information into an object file in perception—as when information reaches it via one of Pylyshyn's FINSTS—are entirely unconscious, at the subpersonal level, and explain facts about perception, something that simply happens to the subject at the personal level. The conditions for entry of information into the perceptual object file require that it be information from the FINST-tracked object, when all is functioning properly, and this concerns an unconscious, computational level of representation.

A subject's phenomenological file on itself is equally constructed below the level of conscious mental action, and below the level of consciousness altogether. The seemings of perception, memory, emotion, and the rest are things that, at the conscious level, just happen to the thinker. Subject-reflexive perceptual seemings whose content concerns *de jure* the thinker himself are just a special case of this more general fact.

An account somewhat analogous to that given here for the first person can also be given for the case of conscious events and states with present-tense contents. The account would refer to a 'now' file on a time. The same applies again to conscious states and events that represent something as happening or being the case *here*. Representations of these states and events would be collected in a 'here' file.

There is a further respect in which a subject's file on itself has a distinctive status. A time can exist without representing itself and having a file on itself, the idea makes no sense. The same applies to places. By contrast, a subject may represent itself, and when it does, it must have a file on itself, a subject file.

A subject's primitive file on itself should not be regarded as integrating representations that already exist as the basis of conscious phenomenal events. You may be aware of your having both the properties of having a pond to your left and of walking straight ahead; but this should not be described as your operating, at the conscious personal level, on two already existing conscious events to somehow make them co-conscious. Your total state of integrated awareness is not a result of your conscious mental action upon some more primitive, already-conscious events and states. Your total state of subjective consciousness is not generated by your conscious mental action at all (though of course it may have mental actions as a component). The subject's file on itself, if it is to contribute to the explanation of subjectivity, must be regarded as operating on representations which

are precursors of the representations that underlie conscious events and states, on pain of misrepresenting consciousness and phenomenology.

In this respect what is integrated in the subject's file on itself is very different from the case in which we have inferential integration that moves at the personal level from the two judgements *I am F* and *I am G* to the conjunctive *I am both F and G*. That inferential integration operates on conscious judgements at the personal level; the subject's file on itself does not. This means that we have to distinguish two kinds of subject file. There is the more primitive one I have recently been discussing, a file which helps to explain how things seem, nonconceptually, to the subject. But there is also, at the level of judgement, something that can still fairly be described as the subject's file on himself, that functions to integrate the contents of conceptual judgements. In cases in which a subject rejects the content of his more primitive, pre-judgemental phenomenology—for example, in the case in which he knows that he is looking at a perfect *trompe l'œil* or hologram—the contents of the more primitive subject file and the personal-level file at the level of judgement will not be in accord with one another. The more primitive one will contain some representation of the content of the illusion. The subject file at the level of judgement will overrule the illusion.

The operation of taking materials from precursors of conscious mental events and integrating them to form the contents of a subject's file on itself may seem in certain respects to resemble the operation of transcendental apperception in Kant's critical philosophy. Some of the resemblances are real, and could be pursued as an independent (and complex) topic. But there are also important differences. Precisely because the operations inserting material into the subject's file on itself are applied to non-conscious precursors of conscious event, they should not be regarded as operating on Kantian intuitions, if those intuitions are regarded as involving even a primitive form of consciousness. Mechanisms that insert something into a subject's file on itself are also open to purely empirical investigation. We could learn more about them by further investigations in empirical psychology.

Despite these differences (and others), there is an undeniable Kantian streak in the position I am outlining. I am concerned to formulate the constitutive conditions of subjecthood. That is a goal that overlaps with Kant's. The goal is partly realized in the same way. The idea of a subject's file on itself, though very differently elaborated in these very different approaches, needs to play a role in both accounts.

The characteristic of a conscious subject-reflexive state or event, that its predicative content is carried through, in central cases, to a subject's file on itself, seems to me to be a resource on which we should draw in addressing two closely related constitutive philosophical questions.

First, we can cite this characteristic in answering the question "What is it for a conscious state or event, such as an experience, to have a content concerning a subject that is also capable of enjoying other conscious states and events?" Without the psychologically real possibility of predicative integration with the contents of other subject-reflexive events and states, it would be hard to answer this question. It is

because the content of an experience enters the subject's file on itself that it can concern a subject capable of being in other mental states too.

Second, I suggest that having a nonconceptual self-representation involves having an object file for oneself into which representations of the predicative contents of subject-reflexive conscious states and events are normally placed. At this level, these three properties are all instantiated:

having an object-file on oneself that takes in the relevant predicative contents;
being capable of subject-reflexive mental events and states; and
having a nonconceptual self-representation.

With these interconnections and their grounds, we begin to move from a mere description of subject-reflexivity, as given in the earlier characterizations, to the beginnings of an explanation of its possibility and its nature.

One apparent attraction of this approach to primitive self-representation is that it gives priority neither to perception, nor to thought, nor to action, nor to sensation, in an account of primitive self-representation. A subject that has perception, but no action-awareness, can meet this condition. So can a Helen Keller. Provided the subject can enjoy states and events with subject-reflexivity, and the predicative content of those states is transmitted to an object-file on the subject, then the conditions for having this primitive nonconceptual representation of the subject are fulfilled.

It is natural to compare this with the attractions of an account of the full-fledged first person concept, according to which its reference is fundamentally determined by the rule that any use of the concept *I* in thinking refers to the thinker, the agent of that thinking. This rule gives priority neither to perceptual input (as Evans did) nor to intention and action (as Brandom does), but rather sees these connections as consequential of a fundamental reference rule for *I* that in itself gives priority to neither (cf. Evans 1982, ch. 7; Brandom 1994, ch. 8, section V.2; Peacocke 2008, ch. 3).

It is reasonable to expect that the subpersonal mental representations involved in enjoying a subject-reflexive state or event contain, or have some functional analogue of, a symbol indicating the self-representation. It could be the presence of this symbol that normally pulls the predicative content of the subject-reflexive representation into the subject file. It then becomes reasonable to ask how this subpersonal symbol, or functional analogue thereof, differs from a subpersonal representation for the first-person concept itself. The distinction between the conceptual and the nonconceptual demands a separate chapter, but there are still some brief answers to this reasonable question.

First, the phenomena and the states and events I have been discussing so far, at the level of perception, memory, and action-awareness, and registration of the contents thereof, can all be present below the level of judgement, a rational and potentially reflective mental activity. The notion of a concept is essentially that of something that features in the contents of judgements. Different current substantive theories of concepts vary in their account of the general form that the individuation of a concept must

take, but they commonly respect this constitutive connection between concepts and judgement. If this connection does exist, then the primitive subject-reflexive states and events that I have been discussing so far need not involve concepts so conceived at all. The integration of the contents of subject-reflexive states, the proper updating of object-files, and the rest, can all be present in the mental states of a being that does not make judgements that are made reasonable, but not forced, by these various non-judgemental states. Similarly, if you hold that critical thinking is essential to possession of a concept, that too is a capacity additional to anything so far cited in this discussion of subjects and subject-reflexive states and events. Finally, many of the contents in which a primitive form of subject representation is involved may be scenario contents of the sort I discussed in *A Study of Concepts*, the wrong sort of content to be conceptual content (cf. Peacocke 1992). In short, we have been operating so far at a level below the kind of rationality and reasons involved in making judgements. In fact, it is precisely because we are below that level that we can use this material in elucidation of the first-person representation that is genuinely conceptual.

There is an argument that we need not two, but three degrees of self-representation to accommodate the phenomena. We can conceive of a creature with perceptions of the world, and whose location in the spatial world changes, but who uses no notion of itself as having a location in the world, neither at the conceptual nor at the nonconceptual level. This creature will use, at the nonconceptual level, and perhaps even at the conceptual level, versions of demonstratives about perceived objects and events. It may represent these objects and events as being at various locations identified in relation to a 'here'. It may represent observable properties and relations as holding between the perceived objects and events. All this can be done without using the notion of these objects as being related to itself. Yet this creature has perceptual states, and there is certainly a subject who has these perceptual states, and experiences the changing objective world. This subject will, like other subjects, have apparent memories of some features of the way the world was at earlier times. But at no point does this subject represent itself as having a location in the spatial world. A fortiori, it does not represent itself as tracing a route through the spatial world as time passes. It has neither states with the content "I am here now", nor states with contents of the form "I was there then".

In imagining this case in more detail, it may help to consider underwater worlds, in which a perceiving subject without an intentionally active body is passively moved around. The underwater subject can still represent things and events that occur in what is in fact its environment, though of course it is not represented by the creature as *its* environment. Someone might doubt that this case is possible on the grounds that the contents of perceptual experience concerning distance and direction are possible only for a creature enjoying bodily agency, and which can act by moving in certain directions and distances, in accordance with its representations of those distances and directions. I suggest, however, that the connections between perception and action are not so tight. I agree that true statements about the representational content of a creature's perceptions, including those about distance and direction, must have

explanatory repercussions for the creature's actions in some possible circumstances. But it does not seem to me that that general constraint implies that spatial perceptual content is possible only where there is spatial action. The kind of creatures envisaged who do not self-represent may act differently according as they are in one part of the world, rather than another, as represented on their map of their environment. The map may distinguish qualitatively similar environments. The creature may act by changing its colour, or its acidity level, or the noise it emits, or its electric charge, accordingly. Perceptions with specific spatial contents will be essential in building up its cognitive map of the world as it changes location over time. Prima facie, all this can be present without the capacity for spatial bodily action on the part of the creature.

It follows that we do need to acknowledge three degrees of involvement of subjects in the representation of the objective world. Since the creature we have just described does not represent itself as an element of the objective spatial world at all, its degree of involvement in such objective representations is zero. This is a limitation on the content of the subject's representations, not the nature of the subject itself. The subject itself really is an element of reality. The states enjoyed by creatures with this Degree 0 involvement of subject-representation in the objective world are states that make possible all the richer degrees of subject-representation.

Are these cases of Degree 0 merely ones in which the reference to the self is implicit? The implicit/explicit distinction is an important one, but it cuts across the distinction in question here. A subject at this Degree 0 may employ a representation with the content *pond to the left*, and that may indeed involve implicit elements. But the implicit reference is, in this case, to a place—the location of what is in fact the perceiving subject—rather than being a reference to a subject or person. The implicit/explicit distinction classifies representations, be they mental representations, or sentences, or utterances, rather than what is represented. The difference between Degree 0 cases and others concerns what is represented, not the vehicle of representation.

The existence of cases in which there is Degree 0 of subject-involvement in the representation of the objective world offers support for an even more radical version of the claim that, way back in *A Study of Concepts*, I called 'the Autonomy Thesis'. The Autonomy Thesis states that a subject might enjoy a set of events and states with nonconceptual contents without possessing a set of genuine concepts at all. This Autonomy Thesis is denied by those—including me in an erroneous past—who think that any content at all, even the nonconceptual kind, must have connections with states with first-personal conceptual contents. The denial of the Autonomy Thesis sometimes rests on a failure to recognize that supposed connections with the first person can be connections, not with the first person concept, but rather with a nonconceptual analogue that has the sort of links we have been discussing with a subject's file on itself, and with subject-reflexive events. But if these Degree 0 cases exist, then the idea of a constitutively autonomous level of nonconceptual content operative in the minds of animals and in our less sophisticated representations understates the position. If Degree 0 cases exist, then the sort of sensitivity to changing

relations as one moves (or is moved) in the spatial world may be a sensitivity that does not require that one represent what is in fact one's current location as one's current location. The location may be represented simply as *here*. Not even a nonconceptual first person is required for representation of an objective spatial world.

The representations involved when there is Degree 0 of subject-representation are in the literal sense, and in respect of the subject itself, 'nonpositional', a word that occurs in the natural translation of Sartre's phrase 'conscience non positionelle de soi'. Sartre also held that 'Toute conscience positionelle d'objet est nécessairement conscience non positionelle de soi' (Sartre 1936, p. 136).[8]

In a popular lecture, Vilayanur Ramachandran writes, "the self, almost by its very nature, is capable of reflection–being aware of itself. A self that's unaware of itself is an oxymoron" (Ramachandran 2003, p. 114).[9] If there are cases of Degree 0, as I have argued, then there are conscious subjects who do not self-represent. A fortiori, such subjects are unaware of themselves. One can restrict the term 'self' stipulatively to subjects who do self-represent, and then drain Ramachandran's claim of any substance. The phrase "almost by its very nature" shows that this was not at all Ramachandran's intention, otherwise the "almost" would not be there. So I am in disagreement with him on this issue.

The next degree of subject involvement in representation of the world, Degree 1, is exhibited by a subject who enjoys states with nonconceptual content that is objective, and which represent the subject has having a location, and as standing in other relations, in the spatial world.

At Degree 2 of subject involvement in the representation of the objective world we have use by the subject of the conceptual first person, as expressed in English by 'I', and the enjoyment of conceptual states that represent the subject as located in the spatial world. Just as the states enjoyed by a creature at Degree 0 make possible the nonconceptual first person states enjoyed at Degree 1, so similarly the states at Degree 1 make possible the conceptual states enjoyed at Degree 2.

If there can be subjects with only Degree 0 of involvement of self-representation in its representations of the objective world, there are some important conclusions to be drawn about the most basic structures underlying the existence of a subject. It is not correct to speak of the existence of a subject file in cases of Degree 0, because the file is meant to include predicates the subject represents as holding of himself; but the subject at Degree 0 does not represent himself at all. Instead of a file at Degree 0, we have just the even more primitive binding of representations (such as the perceptual representation of something heard on the right with something seen straight ahead). These more

[8] I am not, however, at all in agreement with Sartre's other theses about nonpositional consciousness, which seem to involve a form of no-ownership thesis about mental states and events. There is further discussion of this in Part II below.

[9] Published in the US as *A Brief Tour of Human Consciousness* (New York, NY: Pearson, 2004), in which the quoted passage is on p. 97.

primitive states and their binding must exist if there is to be a subject at all. At Degree 1, the binding function and the realization of a simple kind of *de se* representation are intertwined. But the binding and the representational functions can in principle come apart, if cases of Degree 0 are possible. There can be subjects without *de se* representation of the subject by the subject.

The points I have been making about nonconceptual self-representation are neutral on the question of whether subjects must, in some central or fundamental case, be embodied. Those who think that subjects must, in the fundamental and explanatorily central cases, have a body can consistently endorse the legitimacy, interest, and importance of the notion of subject-reflexive events and states. Those theorists could consistently insist that embodied subjects may enjoy subject-reflexive states as characterized here. So could those who deny that subjects must, in some central cases, be embodied. The notion of a subject-reflexive state can by itself serve several radically different ontological views.

II

I turn to the second of the questions identified at the outset, that of the metaphysics of subjects, the nature of the entity referred to in self-representation.

The nature of subjects and the nature of conscious states and events are ontologically interdependent. In one direction, there is a dependence because:

> what makes something a conscious state or event is that there is something it is like for the subject of that state or event to be in that state, or to be the subject of that event.

This is a statement of the classical characterization of Thomas Nagel, and the characterization entails that conscious mental events and states have subjects (cf. Nagel 1974). The characterization does not in itself say anything about what it is for a mental event or state to have a subject. What I recommend is that we look at this intuitive characterization of consciousness as one part of an account of the ontological interdependence of mental events and their subjects. As Frege said, "It seems absurd to us that a pain, a mood, a wish should go around the world without an owner, independently. A sensation is impossible without a sentient being. The inner world presupposes somebody whose inner world it is" (Frege 1977, p. 14). In contrast, "Things of the outer world are on the contrary independent" (ibid.), and do not need any bearer. The claim of ontological interdependence disputes the view Hume formulates when he says of "particular perceptions" that "All these . . . may exist separately, and have no need of anything else to support their existence" (Hume 1739–40, Book 1, Part IV, sect. 6, par. III).

The principle I am defending—that what makes something a conscious state or event is that there is something it is like for the subject of that state or event to be in that state—is a metaphysical principle about the nature of the property of being conscious.

It makes a claim about a certain type of event or state. Using the very helpful apparatus of Kit Fine, we would formulate the claim thus: it is in the nature of the property of being a conscious occurrence that for anything with that property, there is some subject who enjoys (or suffers) that occurrence. In the Appendix to this chapter, I give a formalization of this claim and some further distinctions using Fine's notation (cf. Fine 1995).

There is an equally plausible dependence in the other direction, a dependence of the nature of subjects on conscious states. The dependence is captured in this principle:

> what makes something a subject is that it is capable of being in conscious states and of being the subject of conscious events.

Here, despite some other disagreements with him, I am in agreement with Sartre (or at least with the spirit of his claim) when he writes that consciousness is the subject's dimension of being: 'la conscience', he says, is 'la dimension d'être du sujet' (Sartre 1948, p. 136).

It is also plausible that the property of being a subject is a subject's most fundamental substantive kind. Anything that is a subject is essentially a subject.

Such a conception of ontological interdependence between subjects and the capacity for being in conscious states is not intrinsically a Cartesian conception. It is entirely consistent with this ontological interdependence that both subjects and their conscious states require some material realization.

The ontological interdependence does not, or at least does not obviously, imply that subjects are essentially or fundamentally embodied. The interdependence does not, or at least does not obviously, imply that subjects are essentially or fundamentally living animals. If embodiment is in some fundamental way necessary to being a conscious subject, that necessity would need to be shown by further arguments. So too would the claim of mere contingency of embodiment.

It may be asked why a particular pain or other mental event isn't individuated in the same way as any other particular event, by (perhaps) its particular causes and effects, as in Donald Davidson, or by some other account of particular events (cf. Davidson 1969). I answer that there is no incompatibility between such view of the individuation of particular, token events and the claim I am making of the interdependence of conscious events and their subjects. The claim of interdependence is a claim about types or kinds: what makes something a pain, for example, is that it bears a certain relation to a subject that suffers it. That is something concerning its kind (and no doubt its most fundamental kind), rather than its identity as a particular, for which some other account may be correct. The distinction between the nature of the kind and the individuation of the particular member of the kind applies to events quite generally, whether conscious or not, whether mental or not. Consider, for instance, explosions. What makes an event a member of that kind is that it is, roughly, a flying apart of some object or mass of material caused by forces acting from within that object or mass. That account of the nature of the kind is entirely consistent with there being

some other account of the individuation of any particular explosion occurring at a particular time.

The interdependent conception I am offering is, by contrast, opposed to 'no-ownership' views of mental states and events. On the approach I am advocating, it is not only essential to conscious states and events that they have subjects, but this possession by a subject is also involved in what makes them conscious events and states. It is this notion of a subject, distinct from all three of Cartesian conceptions, from conceptions of essential embodiment, and from the constructed subjects of no-ownership views, that I would argue is also crucial for the philosophical elucidation of such matters as self-consciousness and the nature of first-person thought.

Subjects can sense, perceive, and think. They can act physically and act mentally. These sensings, perceptions, thinkings, and actions are conscious states and events of the subject. Subjects persist through time. Some conscious states are experienced by their subjects as continuing conscious states. Subjects can remember some of their previous conscious states and events; and some subjects are capable of thinking about their future states and events. Many of these mental states and events each have correctness conditions concerning the subject of that very state or event.

Hume famously remarked, "when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* at any time without a perception, and never can observe anything but the perception" (Hume 1739–40, Book I, sect. VI). What is right in Hume's remark does not, as he thought, tell against an ontology of subjects of the sort I am advocating. What Hume's remark highlights is something quite different. It highlights rather the fact that what G. E. Moore called the diaphanous character of conscious events and states applies equally to the subject itself, as it can be given in conscious states and events. Moore described the diaphanous character of consciousness thus: "the moment we try to fix our attention upon consciousness and to see *what*, distinctly, it is, it seems to vanish: it seems as if we had before us a mere emptiness. When we try to introspect the sensation of blue, all we can see is the blue: the other element is as if it were diaphanous" (Moore 1903, p. 41). In conscious states and events, the subject is diaphanous in the way in which Moore rightly says the consciousness or awareness is diaphonous. In enjoying an experience or any other conscious event, we can equally say of the subject of the conscious event that "the moment we try to fix our attention upon the subject and to see *what*, distinctly, it is, it seems to vanish". The correct response to Hume's position is that he, the subject, no more needs to be an object of awareness and attention to exist and to be involved in his current conscious states and events than consciousness itself needs to be an object of awareness to be involved in his current conscious states and events. The ontological status of the subject is no more impugned by the phenomenon of diaphanousness than is consciousness itself.

Hume did not draw the same conclusion about consciousness itself that he drew about the subject of experience. Almost everyone would regard it as a

reductio ad absurdum of his form of reasoning if he were to have done so. So we have to avoid a double standard that accords to subjects a treatment that it does not apply to consciousness. Hume's line of thought gives no reason to dismiss an ontology of subjects that would not apparently equally apply to consciousness itself. Suppose we write (very crudely, and ignoring distinctions important for other purposes) the basic form of attribution of a conscious state as

(1) Subject *x* enjoys event *e* of conscious kind *K* with intentional content *C*.

What we are aware of in enjoying a conscious state or event is what is given in the content *C* in the mental event or state of kind *K*. The four elements—*x*, the subject of the awareness, the conscious state or event *e*, the intentional content *C*, and what at the level of reference is given in *C*—are equally and essentially involved in this state of affairs. Neither the subject nor the consciousness is involved by being something on which the subject can fix his attention. The consciousness is of what is given in a certain content, in a state of kind *K*. It is equally a consciousness belonging to the subject.

I do not mean to imply that whenever a subject is in a conscious state, the subject has some awareness of himself as being in that state. That would collapse the distinction between consciousness and self-consciousness. We are concerned here with a subject's having a mental event or being in a mental state, rather than with the subject representing something as being his own in enjoying that event or state. The thesis is only that in any case of a conscious event or state, there is a subject who enjoys that state or event, and that it is a nonsequitur to argue from the fact that one cannot attend to or be aware of the subject in a certain way to the conclusion that the subject is not so involved. But I do add that even when a subject is aware of his being in a conscious state, as when you are aware of your seeing a red traffic light, you need not be given to yourself in a way which allows you to attend to yourself.

Still, it may be objected, are you not attending to yourself when you look at yourself in a mirror, or when you look at your hand to examine a cut? Isn't there an entirely unobjectionable and commonplace phenomenon of attending to yourself? I agree that there is; but I also want to distinguish. There is a distinction, in examples of attending, between what we can call *derivative* and *original* attention. When you attend to yourself by attending to yourself in a mirror, you attend to yourself by attending to something not given in a first person way, but rather to something given as your body. The same applies when you are attending to your hand. In these cases, you are attending to yourself by attending to something given in a way distinct from the first person way. By contrast, when you attend to the road in driving, or to the traffic light, you do not do that by attending to something given as anything other than the road, or as the traffic light. When you attend to one of your throbbing pains, you do not do that by attending to something given in some way other than as that very pain. When you attend to the road, or the pain, or the light, these are cases of original attention. The thesis that you cannot attend to yourself should more properly be formulated as the

thesis that a case of attending to yourself cannot be a case of original attention. It can be only a case of derivative attention.

Hume was surely well aware that he could attend to his own body. It is very unlikely that he would have taken this obvious fact as a counterexample to what he meant when he wrote that he could never catch himself when he enters most intimately into himself. If this is right, then Hume could have used the notion of original attention. He could have said: any example of attending to yourself is an example of derivative attention. In my view, Hume would be right in elaborating his position in that way. He was wrong to presume that everything that is essentially involved in consciousness must be something to which one can attend in a case of original attention.

Attending is a relation between a subject and an object, but when we speak, as I have, of "attending to something given in a different way", the basic relation from which that concept is built is this: attending to *x* as given in way *m*. Consider the example, familiar from discussions of identity, of an aircraft carrier seen through one window, and an aircraft carrier seen through a second window, where in fact the same carrier is seen through both windows. It makes sense to say that someone is attending to the aircraft carrier as seen in the window to the left, as opposed to the attending to the aircraft carrier as seen in the window to the right. When we use the intuitive locution "attending to one thing by attending to something given in a different way", this means: attending to an object given in one way *m* by attending to an object given in a distinct *way m′*. So attending to an object given in a different way does not necessarily mean attending to a different object (it would not be a different object in the example of the aircraft carrier). What matters is that the ways are distinct, not that the objects given in the two ways are distinct. The ways *that person* and *I* are certainly distinct. That suffices for correct application of the distinction between original and derivative attention, and its use in discussing Hume's point. The distinction between the two cases of attention leaves open the question of whether the entity given in the first person way is the same thing as the entity given as *that person*. It follows that these points can be endorsed even by someone who thinks that you are identical with your body, or holds that you are in some way individuated by your body.

The distinction between derivative and original attention takes us a certain distance. We should, however, demand much more by way of explanation of why some cases of attention fall on one side of the distinction, and others fall differently. Why is there apparently no such thing as attending to oneself originally? An immediate answer is that to be an object of attention, the object or event must be given in perception, sensation, or perhaps in certain kinds of sensory imagination, and to be given in one of those ways is not to be given as oneself. Again, this seems fine as far as it goes. The explanatory question, however, is just pushed back. Why is there no such thing as being given as oneself in perception, sensation, or certain kinds of sensory imagination? To ask the question in compressed form: why can't the subject be an object of perception?

In the previous section, I suggested that the first person nonconceptual notion is individuated at least in part, if not wholly, by the condition that in any mental state,

event or process in whose content it features, it refers *de jure* to the subject (the possessor) of that state, event, or process. To be given in perception, sensation, or sensory imagination is always and necessarily different from being given as the subject of a state, event or process. There is such a thing as being given as the subject of an event. If, for instance, you are aware that you are seeing the phone, or are aware that you are in pain, in your awareness you are given as the subject of the seeing, and of the pain, respectively. But in neither case, nor in any other, are you given as yourself in the scene perceived, or in the sensed state of affairs. One cannot be given as the subject of a mental event in the state of affairs as perceptually presented in the event.

The source of this incompatibility is the very nature of the nonconceptual notions involved. To be given as the subject of an event is one thing; to be given in a perceptual, sensational, or imagistic way is a different thing. It is neither a contingent nor a merely a posteriori fact that Hume could not find himself in any of his impressions. Hume says, no doubt ironically, that another person "may, perhaps, perceive something simple and continu'd, which he calls himself; tho' I am certain there is no such principle in me" (Hume 1739–40, Book I, sect. IV, part 6). We need to be explicit about what would justify the irony, to establish that Hume's observation is not merely one about some individual's empirical psychology. I suggest that an explicit statement needs to appeal to the different natures of the first person notion and perceptual and other notions that are not first personal. This explanation of why there is no such thing as attending originally to oneself appeals not to the nature of the objects of attention, but to the distinctions between the various ways in which things are given when one is perceiving and attending.

It may be helpful here to compare the first person in this respect with the nonconceptual present tense notion *now*. It displays some relevantly parallel phenomena. When it features in the content of any perception or other mental state, event, or process, *now* refers to the time of occurrence of that state, event, or process. (It does so non-descriptively, of course, just as the first person is non-descriptive.) Now a perception may also involve the representation of other events, as for instance when a television producer sees many screens simultaneously. Some of the events on the screens may be of past events, some may be of present events. But none of them is given, in the experience, as happening now (as opposed to the displayed images being given as occurring now, which of course they are). The time of occurrence of an event represented on one of the screens is given as "the time of this screened event", if we may allow demonstrative reference to events presented on TV. That is, in its nature, a different way of being given than as the time of the perception in which the TV screens are perceived. Correspondingly, at a higher, conceptual level, "that event (as presented on a particular TV screen) is happening now" is always potentially informative.

Because nothing given in perception is thereby given as the subject, and because this holds whatever perception may be in question, a temptation may exist to say that the subject is not in the world, or that the subject is merely the limit of the world. Neither proposition follows. The fact that the subject is not given in perception as being in the

world does not imply that it is not in the world. The fact about the impossibility of perceiving something as the subject is a fact about the way in which something is given. This should not be confused with a fact about the entity so given.

Wittgenstein did succumb to the temptation when, in a much-discussed passage of the *Tractatus*, he wrote:

> 5.632 The subject does not belong to the world: rather, it is a limit of the world.
>
> 5.633 Where *in* the world is a metaphysical subject to be found?
>
> You will say that this is exactly like the case of the eye and the visual field. But really you do *not* see the eye. (cf. Wittgenstein 1921)

Anything one perceives or senses would be in the world, mental or non-mental. This and the preceding section supply a reading of Wittgenstein's remarks as based on a genuine insight (also in Schopenhauer) about a distinctive way of representing themselves that is available to subjects. But really nothing follows from these insights about subjects not being in the world. This is one of many points at which the insights of the *Tractatus* would have been formulated very differently if Wittgenstein had made use of Frege's distinction between sense and reference, and had applied it at every level of representation, both conceptual and nonconceptual.

On the position I am developing here, subjects and conscious events are ontologically coeval. Specifying the nature of either one involves mention of the other. So mental events already involve subjects, and not just as a way of talking. The position I am advocating is not consistent with the view that the existence of a subject consists in the existence of various mental events and other entities. I am taking it that in that neo-Humean view, 'consists in' is supposed to be an asymmetrical relation. If 'consists in' is an asymmetrical relation, my position is incompatible with the neo-Humean thesis that the existence of a subject consists in the existence of mental events and other entities that allegedly do not involve subjects.

Derek Parfit formulates his famous view of these matters as the thesis that a subject is not a 'separately existing entity, distinct from a brain and a body, and a series of physical and mental events' (Parfit 1987, p. 223). Suppose '*A* exists separately from the *B*s' is understood to imply 'The existence of *A* is not settled by or determined by the existence of *B*s'. Then of course I would have to agree that in that sense a subject of experience is not an entity existing separately from mental events. But this agreement is not an agreement to a form of reductionism or constructionism about subjects. A subject does not 'separately exist' on this understanding of the phrase 'separate existence', because mental events, on the present view, already involve subjects. So the existence of the subject *is* in one sense settled by the existence of mental events and states.

It follows that the formulation in the quotation I gave from Parfit does not fully capture his intention, the intention to formulate a species of moderate reductionism about subjects of experience. The issue of reductionism should really be formulated in terms of individuation and the ontological priority, or otherwise, of mental events and

states vis à vis the subjects that enjoy them. The issue at stake in the reductionism in which Parfit is interested is that of whether mental events and states are ontologically prior to the subjects who enjoy them.

Since the fundamental issues are metaphysical, it also follows that we should be careful not to formulate the issues in terms of what is describable without explicitly mentioning subjects. We can conceive of what Derek Parfit calls an 'impersonal' description of mental events, bodily events, and physical objects and events, and their properties and relations, a description that does not explicitly mention subjects (Parfit 1987, p. 225). This does not at all mean that there is no commitment in such a description to the existence of subjects. As a comparison: we cannot establish relationism about space simply by giving a description of everything we want to say in terms of spatial relations between material things and events. If a case in metaphysics can be made that spatial relations between material things themselves consist in relations between the places at which the things are located, then the availability of the place-free description fails to establish relationism about space. The same applies to the impersonal description. If the events and the properties of events mentioned in the description cannot be elucidated without mentioning subjects, then all we have in such a description is something that does not make explicit the full ontology to which it is committed.

That any such philosophical elucidation of mental events and their properties, in particular the property of being a conscious event, must involve subjects is precisely what I have been arguing. For someone who thinks that talk of subjects is a mere *façon de parler*, there must be some replacement for the characterization of a conscious state as one such that there is something it is like to be in it for the subject of that state. I have no idea what this replacement could be.

There is, not at all surprisingly, an analogous problem for a view of action that parallels Hume's view about conscious events. We can imagine a theorist who holds that there is no subject who is the agent of actions, 'existing separately' from the actions themselves and other events and entities. Actually we do not have to imagine such a theorist, for Nietzsche held precisely this view of agents. He wrote, "For just as the popular mind separates the lightning from the flash and takes the latter for an *action*, for the operation of a subject called lightning, so popular morality also separates strength from expression of strength, as if there were a neutral substratum behind the strong man, which was *free* to express strength or not to do so. But there is no such substratum; there is no 'being' behind doing, effecting, becoming; 'the doer' is merely a fiction added to the deed—the deed is everything" (Nietzsche 1887, p. 481). A few lines later he adds, "our entire science still lies under the misleading influence of language and has not disposed of that little changeling, the 'subject' (the atom, for example, is such a changeling, as is the Kantian 'thing-in-itself')".

The problem for this view is that what makes something an action is that some subject does it; just as what makes something a conscious mental event is that there's a subject for which there is something it's like to experience (or enjoy) the event.

I equally have no idea what Nietzsche's replacement account of what makes an event an action could possibly be.

I now turn to some wider issues in the ontology of subjects. To say that the ontology of subjects and the ontology of mental states and events are interdependent is not at all to imply that subjects are immaterial things. On the contrary, such interdependence may imply just the opposite. If various aspects of the mental states and events with which the ontology of subjects is interdependent themselves require certain relations to the nonmental world, that fact will have implications for the material status of subjects too. Just that seems to be the case. The mental states and events have multifarious constitutive connections with the material world. A sensation of pain must be experienced as in some part of some apparent body. What it is for it to be experienced as in some part of an apparent body cannot be elucidated independently of relation of the pain-event to such a body part, when the subject has a body with such a part. What it is for a visual experience to represent something as being a certain distance and direction from the subject cannot be elucidated independently of the causal powers of things at certain distances and directions from the subject, when the subject is properly connected to the world—and so forth, for myriad other examples.

If something is in states that are individuated in part by their relations to material objects, properties, and relations, it seems as a matter of general metaphysics that it must have some realization in material objects and events. This is true of domains at some distance from our present subject matter. It is true of a financial institution such as a bank, of which it is constitutive that it can stand in loan, credit, and debit relations. The transactions of a bank must be materially realized, even those of an internet bank. We have, apparently, no conception of how something could stand in these material relations without itself having a material realization. The same seems to me to be true of subjects of conscious states and events. So Cartesian immateriality not only need not be, but also should not be, any part of the conception of subjects for which I am arguing. For this reason, I also think that the conception of subjects for which I am arguing is consistent with Kripke's famous views on the essence of origin for persons and, by extension, subjects (cf. Kripke 1980).

The identity of a subject over time is something to be explained partly in material terms. That a subject has some material realization plausibly follows from these considerations about the material realization of states that are themselves individuated by their relations to material things and events. More specifically, I suggest that the identity of a subject over time consists in the identity of the apparatus that integrates states and events in such a way that a single subject has, or may have, perceptions, sensations, thoughts, action-awareness, and the rest, both at a time and over time. What matters is the identity of the integrating apparatus, not the identity of the particular pieces of apparatus whose states are integrated to yield states of the subject. Your perceptual apparatus may be entirely replaced consistently with your continued existence, provided that the states produced by the new apparatus are properly integrated by some continuing apparatus with your other conscious states and events.

Like any other material object, the matter constituting the integrating apparatus may also change over time. This partial account of identity of subjects over time provides a connection between what makes something a subject—its ability to enjoy a range of kinds of conscious states and events—and the role of the material integrating apparatus in which the identity of a given subject consists.

Some thinkers hold that a subject does not persist if all its memories and beliefs are destroyed. For these thinkers, continued existence of the integrating apparatus in which a subject's existence is partially realized does not amount to continued existence of the subject itself. I myself do not share this intuition about the case. But for those who do, all that matters for present purposes is that identity of integrating apparatus is a necessary condition for continued existence of a subject, even if it is not sufficient. Further conditions can be added to reach sufficiency if such conditions are thought to be needed.

The integrating apparatus provides the realization of the subject's file on itself, the file discussed in the preceding section on primitive self-representation. This is one of several connections between the theses on representation in that section and the metaphysics of subjects outlined here.

There is also a connection between the metaphysical interdependence of conscious events and subjects, on the one hand, and the correct description of the operation of the subject's file on itself. I emphasized that the file should not be regarded as collecting together predicates in the contents of events that are already conscious. If they were already conscious, it certainly seems that there could be conscious events that could exist both temporally and ontologically prior to there being a subject who enjoys them (since the subject file is the causal basis of the subject's being in a number of co-conscious states). On the present view, the correct conception of the subject file and the ontological interdependence of conscious states and events go hand-in-hand.

Kant famously, and as it seems to me correctly, objected to one conception of conscious subjects for the reason that it fails to distinguish between the existence of one continuing subject on the one hand, and a succession of distinct shorter-lived subjects which pass memories of their states on to their immediate successors, which in turn are ignorant of the change in identity (cf. Kant 1787, p. 423). Kant's point need not be construed as verificationist. It can be regarded as a constitutive challenge to conception of subjects in question. If the ontology of subjects is legitimate, what is the difference between being sensitive to genuine identity over time, as opposed to apparent identity realized in a succession of subjects? The question is analogous to nonverificationist objections to classical Newtonian absolute space (cf. Peacocke 1988).

I think the constitutive question is well posed against any position that makes genuine identity consist in seeming-identity. But the present account is not such a position. Genuine identity of subject consists in real identity of the material, integrating apparatus in which a subject is realized. There could in extreme, distant counterfactual cases be transmissions of memories through a succession of distinct underlying physical pieces of integrating apparatus. That really would involve an illusion of identity over

time on the part of the subjects in the later parts of the series. So apparent identity by no means ensures genuine identity on the present view. It meets Kant's justified demand.

III

The agenda of this chapter is just a first step towards a fuller account of subjects of consciousness and first person representation. A fuller account must deal with a variety of richer and important notions of self-consciousness that go beyond mere first person representation, and it must account for the epistemological and psychological significance of those notions of self-consciousness.

That fuller account should also apply the conception I have developed here to the classical dispute between Descartes and Kant on rational psychology, and to more recent treatments of the first person and the self. It is striking that, in contrast to many other recent theories of subjects and the first person, a subject's body seemingly plays no fundamental role in the account I have offered of subjects and of first person representation. Yet we still have to give a constitutive account of the distinctive phenomenon of embodiment, and we have to explain its ramifications. This all has to be done whilst also respecting the important role the first person plays in the individuation of many notions and concepts. We need a general reorientation and restructuring of our thought about the self and self-representation.[10]

★★★

Appendix: Natures, Properties, Occurrences, and Subjects

I said in Section I that it is in the nature of the property of being a conscious occurrence that for anything with that property, there is some subject who enjoys (or suffers) that occurrence. We can formalize this claim, and make its commitments more explicit, by using the notation in Kit Fine's paper 'The Logic of Essence'. Fine there introduces the notation $\Box_F A$, which is to be read as: A is true in virtue of the nature of the objects which F. Let C be the property of being a conscious event or occurrence. Then the claim I am making about that property is that:

(i) $\Box_{\lambda P(P=C)} \forall x(\, Cx \rightarrow \exists y(y \text{ is the subject who enjoys (suffers) } x))$.

[10] Material overlapping with this chapter was presented in 2008 at Syracuse University, at a Special Lecture at Oxford University, and at the Conference 'Self and Self-Knowledge' at the Institute of Philosophy in London. Versions were also presented to my seminars at Columbia University and University College London, and at the Brown University conference on perception in 2009. I thank Andre Gallois, John Hawthorne, James Higginbotham, Patricia Kitcher, Rory Madden, Michael Martin, Lucy O'Brien, Ian Rumfitt, Paul Snowdon, and Ralph Walker for comments; and Ned Block and Jesse Prinz for their extended remarks on a presentation of this material at the NYU Language and Mind Seminar in Spring 2010. Some of this material was also covered in the first of my 'Context and Content' Lectures at the Institut Jean Nicod, Ecole Normale Supérieure in Paris in October 2010. The text has benefited from my discussions there with Jérôme Dokic, Pierre Jacob, Joëlle Proust, François Recanati, and Georges Rey. The penultimate draft was improved by comments from Antonia Peacocke.

For a range of individual (nonsocial) mental events, it may be in the nature of any such event *e* that it has its particular subject essentially. That further claim, using Fine's notation, is that for any such event *e*,

(ii) If *s* is the subject of *e*, then $\Box_{\lambda x(x=e)}$ (*s* is the subject of *e*).

The proposition with the converse relation embedded in the 'nature of' operator is false. It is not in the nature of *s* that *e* be one of its mental events. The conscious subject *s* could have existed and had a different mental life, one in which *e* does not exist at all. So,

(iii) If *s* is the subject of *e*, then $\sim\Box_{\lambda x(x=s)}$ (*s* is the subject of *e*).

References

Bayne, T. 2004 "Self-Consciousness and the Unity of Consciousness", *The Monist* 87/2, 219–236.

Brandom, R. 1994 *Making It Explicit: Reasoning, Representing, and Discursive Commitment*, Cambridge (Mass.), Harvard University Press.

Davidson, D. 1969 "The Individuation of Events", in N. Resche and D. Reidel (eds.) *Essays in honor of Carl G. Hempel*, Dordrecht, Reidel, 1969, 216–34; repr. in D. Davidson *Essays on Action and Events*, Oxford, Oxford University Press, 2001, 163–81.

Evans, G. 1982 *The Varieties of Reference*, Oxford, Oxford University Press.

Fine, K. 1995 "The Logic of Essence", *Journal of Philosophical Logic* 24, 241–73.

Frege, G. 1977 "Thoughts", in G. Frege (ed.) *Logical Investigations*, tr. P. Geach, R. Stoothof, Oxford, Blackwell, 1–30.

Grice, H. P. 1969 "Vacuous Names", in D. Davidson and J. Hintikka (eds.) *Words and Objections: Essays on the Philosophy of W. V. Quine*, Dordrecht, Reidel, 118–45.

Hume, D. 1739–40 [2000] *A Treatise of Human Nature*, D. F. Norton and M. J. Norton (eds.), Oxford, Oxford University Press.

Jeshion, R. 2002 "Acquaintanceless De Re Belief", in J. Campbell, M. O'Rourke, and D. Shier (eds.) *Meaning and Truth: Investigations in Philosophical Semantics*, New York, Seven Bridges Press, 53–78.

Kahneman, D., Treisman A., and Gibbs B. J. 1992 "The reviewing of object files: Object-specific integration of information", *Cognitive Psychology* 24, 175–219.

Kant, I. 1787 [1998] *Critique of Pure Reason*, tr. P. Guyer and A. Wood, Cambridge, Cambridge University Press.

Kripke, S. 1980 *Naming and Necessity*, Oxford, Basil Blackwell.

Lockwood, M. 1971 "Identity and Reference", in M. Munitz (ed.) *Identity and Individuation*, New York, NYU Press, 1971, 199–211.

McDowell, J. 1984 "*De Re* Senses", *The Philosophical Quarterly* 34, 283–94.

Mach, E. 1914 *The Analysis of Sensations*, tr. C. Williams and S. Waterlow, Chicago (Ill.), Open Court.

Moore, G. E. 1903 "The Refutation of Idealism", *Mind* 12/48, 433–53; repr. in T. Baldwin (ed.) 1993 *G. E. Moore: Selected Writings*, London, Routledge, 23–44.

Nagel, T. 1974 "What Is It Like to be a Bat?", *Philosophical Review* 83, 435–50.

Nietzsche, F. 1887 [2000] "The Genealogy of Morals", *Basic Writings of Nietzsche*, tr. W. Kaufmann, New York, Random House.

Parfit, D. 1987 *Reasons and Persons*, Oxford, Oxford University Press.

Peacocke, C. 1981 "Demonstrative Thought and Psychological Explanation", *Synthese* 49, 187–217.

—— 1988 "The Limits of Intelligibility: A Post-Verificationist Proposal", *Philosophical Review* 97, 463–96.

—— 1992 *A Study of Concepts*, Cambridge (Mass.), MIT Press.

—— 2008 *Truly Understood*, Oxford, Oxford University Press.

Prior, A. 2003 *Papers on Time and Tense* (New Edition) P. Hasle, P. Øhrstrøm, T. Braüner, and J. Copeland (eds.), Oxford, Oxford University Press.

Pylyshyn, Z. 2007 *Things and Places: How the Mind Connects with the World*, Cambridge (Mass.), MIT Press.

Ramachandran V. S. 2003 *The Emerging Mind: The Reith Lectures 2003*, London, Profile Books.

Sartre, J. P. 1936 [2003] *La Transcendance de l'Ego et autres textes phénoménologiques*, V. de Coorebyter (ed.), Paris, Librairie Philosophique J. Vrin.

—— 1948 "Conscience et connaissance de soi", repr. in de Coorebyter (ed.).

Shepard R. and Chipman S. 1970 "Second-Order Isomorphism of Internal Representations: Shapes of States", *Cognitive Psychology* 1, 1–17.

Strawson, P. 1974 *Subject and Predicate in Logic and Grammar*, London, Methuen.

Wittgenstein, L. 1921 [2001] *Tractatus Logico-Philosophicus*, tr. D. Pears and B. McGuinness, London, Routledge.