

José Luis Bermúdez

*Selves, Bodies,
and Self-Reference*

*Reflections on Jonathan Lowe's
Non-Cartesian Dualism*

Abstract: *This paper critically evaluates Jonathan Lowe's arguments for his non-Cartesian substance dualism (henceforth: NCSD). Sections 1 and 2 set out the principal claims of NCSD. The unity argument proposed in Lowe (2008) is discussed in Section 3. Throughout his career Lowe offered spirited attacks on reductionism about the self. Section 4 evaluates the anti-reductionist argument that Lowe offers in Subjects of Experience, an argument based on the individuation of mental events. Lowe (1993) offers an inventive proposal that the semantic distinction between direct and indirect reference delineates the metaphysical boundary between self and world, and uses this as a further argument against reductionism about the self. This proposal is discussed in Sections 5 and 6.*

With the untimely death of Jonathan Lowe in January 2014, analytic philosophy lost one of its most productive and innovative thinkers. Lowe was a very wide-ranging philosopher who made significant contributions to philosophical logic, metaphysics, and the philosophy of mind. He was never one to follow the shifting tides of philosophical fashion. In fact, many of us will remember him fondly for his spirited defence of a number of positions that many (most?) philosophers would deem to be self-evidently false. A fierce opponent of physicalism in both its reductionist and non-reductionist forms, Lowe vigorously propounded his own version of substantial dualism, combined

Correspondence:
Email: jbermudez@tamu.edu

with a robust psychophysical interactionism. As if this were not enough, he offered ingenious defences of the sense datum theory of perception, the volitional theory of action, and the ideational view of language. Proponents of these currently unfashionable positions have a tendency to argue for them indirectly by attacking the competing positions. One of the many merits of Lowe's philosophical style and temperament is that he offered a series of clearly articulated and carefully developed arguments in support of almost all his major theoretical claims, in addition to trenchant criticisms of competing views.

1. Lowe's Non-Cartesian Substance Dualism

According to NCSA selves are enduring and irreducible substances that are non-identical with their bodies. But Lowe's version of this claim is not Descartes'. He emphasizes that selves are subjects of experience, rather than immaterial entities. What distinguishes selves from their bodies is not a radical ontological distinction between material and immaterial substances, but rather more prosaic considerations of persistence and causation. Selves have different persistence conditions from bodies and different causal powers.

As he explicitly states (Lowe, 1996, p. 35), NCSA is intended to do justice jointly and severally to the Cartesian view that persons are a distinctive type of entity, to the Lockean view that this type of entity is essentially a psychological type, and to the Aristotelian view that persons are not essentially immaterial. For Lowe, selves are psychological substances (that is, they require suitably unified successions of psychological states) — that is a point of agreement with Locke. And (a point of agreement with Descartes) selves are simple — that is, they are substances that have no substantial parts — and in particular no physical substantial parts. But Lowe denies (with Aristotle) that selves can be separated from their bodies. In fact, he allows that subjects of experience can be the bearers not just of mental properties but also of those physical properties that they have in virtue of possessing a body that possesses those properties (Lowe, 2008, p. 95).¹

The first difficulty for the sympathetic reader is clarifying the precise metaphysical doctrine being defended. There is, for example, an

¹ Following Lowe (see 1996, p. 1), I am using 'self', 'person', and 'subject of experience' interchangeably.

equivocation in Lowe's presentation of Aristotle's insight that persons are not essentially immaterial. Is this supposed to entail that persons are essentially embodied? We are told that persons cannot be separated from their bodies 'except perhaps conceptually or purely in imagination' (Lowe, 1996, p. 35), and this seems to be reinforced by the stress on the self perceiving and acting through a particular body. But what is the content and force of the claim that self and body are distinct? Like Duns Scotus, Lowe believes that two things can fail to be really distinct and yet be more than conceptually distinct.² Two points secure for Lowe what Scotus would call a 'formal' distinction. First, even though it seems to be necessary that any given self have some body, the connection between any given self and its body is purely contingent — as he puts it, 'this body is only *contingently mine*' (*ibid.*, p. 7, emphasis in original). Second, the self and its body have different persistence conditions (*ibid.*, p. 34).

One might reasonably wonder how either of these two claims could be motivated without begging the question. The first point could neutrally be put as the intuitively plausible claim that I could have had physical characteristics other than the ones that I do have. But this intuition can be cashed out in two ways — as saying either that my body could have been different or that I could have had a different body. Only the second version suggests a contingent connection between self and body, but there is nothing in our initial neutral intuition to suggest the second version. In order to motivate the idea that I could have had a different body, Lowe needs to appeal to thought experiments in which, for example, as Locke suggests, a Prince and a Cobbler swap bodies. These thought experiments would be accepted by many reductionists about the self, but quite plainly would fail to convince anyone who thinks that selves are human animals.

Nor is it clear that Lowe's own theory allows him to assert that the self and the body have different persistence conditions, since his doctrine of the simplicity of the substantial self entails that 'we lack any proper grasp of what would constitute the ceasing-to-be of a self' (*ibid.*, p. 43). We do not, Lowe thinks, have criteria of identity for simple substances, because simple substances lack substantial parts and all criteria of identity rest upon the holding of certain relationships among a substance's substantial parts. Of course, Lowe does not think

² For discussion of Scotus's views in this area see Adams (1982).

that selves are immortal. So there must be circumstances in which they cease to be. We just don't know what those circumstances are. But then, by the same token, we cannot responsibly say that the circumstances in which selves cease to be are different from the circumstances in which their respective bodies cease to be. We have to be agnostic.

Lowe's argument would be better put, I think, as the inference that the self cannot be identical with the body from the principle of the non-identity of discernibles and the lemma that the body but not the self has determinate persistence conditions. This, however, places the full weight of the argument onto the claim that the self is a simple substance, which in turn is derived from the thought that the self cannot have substantial parts since such parts could only be the parts of the body, which would have the result that the body and the self would share all their parts (and hence, on standard mereological assumptions, be the same thing). Of course, one man's *modus tollens* is another man's *modus ponens* and some philosophers will simply conclude from all this that the body and the self are necessarily identical.

In any event, the reasoning here does seem to be moving in a rather tight circle. If the self is a simple substance, then we should certainly grant Lowe that the self is distinct from its body, since bodies do have substantial parts. But why think that the self is a simple substance? Because otherwise it would end up being identical to the body, since its substantial parts would be bodily parts. Wait — that's the very conclusion we are trying to establish!

In Section 3 we will turn to Lowe's 'master argument' for NCSD — the unity argument — which offers a different way of motivating the non-identity of self and body. Before doing so we turn in Section 2 to the second aspect of NCSD, *viz.* the idea that selves have distinctive causal powers.

2. Lowe on Mental Causation and Personal Agency

In Chapter 3 of *Subjects of Experience* Lowe considers and eventually rejects three interactionist proposals to reconcile the notion of independent mental causal powers with the conservation laws of physics. His main objection is that they all display a misplaced allegiance to the Cartesian principle that the mind can only exert causal influence on the body by setting matter in motion. Lowe thinks that the Cartesian principle is incompatible with the structure of the causal chains involved in deliberative action. He models these causal

chains as fractally structured trees converging on a bodily movement. There is no single linear causal chain. Instead, going backwards from every event there are many different chains, splitting and separating off from each other like branches on a tree. Unlike real trees, though, these causal trees have no identifiable tips (at which the mind might set the whole causal process in motion) because the causal ancestry of any bodily movement merges into the prior causal history of the whole brain (Lowe, 1996, p. 65, n. 15). Instead, Lowe suggests, we need to replace the dominant conception of mental causation with a model of mental events as indirect rather than direct causes of bodily movements. Mental events do not directly cause all (or any) of the physical events in the fractal tree convergent upon a bodily movement. Rather, the causal efficacy of a given mental event lies in the fact that such a convergent fractal tree exists at all (*ibid.*, p. 67). We need to appeal to mental events because the existence of such a convergence seems inexplicable in purely physical terms.

Here is Lowe's example in *Personal Agency*. Suppose that someone deliberately raises their arm in order to ask a question in a lecture. Tracing the purely bodily causes of the arm movement back in time would yield a highly complex branching structure in the brain and nervous system, but introspectively things seem very different.

Many of the neural events concerned will be widely distributed across fairly large areas of the motor cortex and have no single focus anywhere, with the causal chains to which they belong possessing no distinct *beginnings*. And yet, intuitively, the agent's mental act of decision or choice to move the arm would seem, from an introspective point of view, to be a singular and unitary occurrence which somehow initiated his or her action of raising the arm. (Lowe, 2008, p. 102)

Here is how Lowe proposes to reconcile the tension.

First of all, the act of choice is attributed to the *person* whereas the neural events are attributed to parts of the person's *body*: and a person and his body are, according to this conception of ourselves, *distinct* things, even if they are not *separable* things. Moreover, the act of choice *causally explains* the bodily movement — the upward movement of the arm — in a different way from the way in which the neural events explain it. The neural events explain why the arm moved *in the particular way* that it did — at such-and-such a speed and in such-and-such a direction at a certain precise time. By contrast, the act of choice explains why a movement of that general kind — in this case, a rising of the agent's arm — occurred around the time it did. It did so because shortly beforehand the agent decided to raise that arm. (*ibid.*, p. 102)

These two causal explanations differ counterfactually. Had the decision not taken place there would not have been an arm movement of that kind at all. And yet, Lowe claims, there is no set of neural events in the bodily antecedents of the arm-raising of which it is true to say that had those events not occurred there would not have been an arm movement of that kind at all. From which he concludes, first, that the decision cannot be identified with any set of neural events, and, second, that the decision is an instance of a fundamentally different type of causation from that operative at the neural level.

Lowe has done philosophers a service by reminding us how impoverished and implausible the causal models which inform contemporary discussion of mental causation are. He may well be right that once we take causal chains to be fractal trees rather than linear successions of billiard-ball-like interactions it becomes completely inappropriate to look for a single initiating cause of a bodily movement that can then be either caused by or identified with a mental event. As he readily admits, however, we really need a positive account of how mental events can have the indirect causal powers that Lowe claims they have. Unfortunately any such account runs into a significant dilemma.

The dilemma comes when we ask whether there is any causal contact that would allow the mental act of choosing to intervene in the branching neural tree. Suppose, first, that there is such an intervention. If there is then there must be an earliest such causal contact, and then Lowe needs to explain why this does not qualify as exactly the type of initiating cause that he is trying to move away from. If the act of choice plays a causal role within the branching neural tree then Lowe seems to lose the distinction between, on the one hand, the neural events that (causally) explain the precise contours of the arm movement and, on the other, the mental event that (causally) explains why there is an arm movement of that type (a signalling).

We can appreciate the problem by looking at the subtle model of mental causation that he offers in Section 3.4 of *Personal Agency*. It is easier to see what is going on in Lowe's own diagram, reproduced as Figure 1.

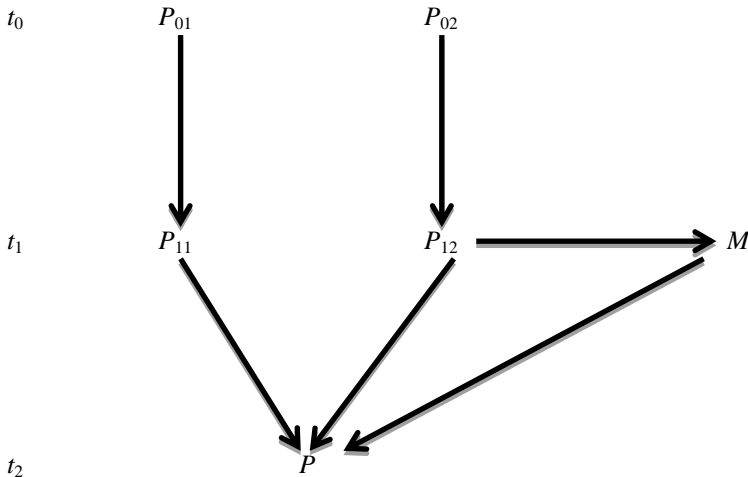


Figure 1. Lowe's model of mental causation (from 2008, p. 71). Note that mental event M is simultaneous with physical events P_{11} and P_{12} .

Lowe's suggestion is that at time t_1 there are two physical events P_{11} and P_{12} , each of which has a purely physical causal history. Those two events are causally sufficient for event P . But P_{11} and P_{12} are only causally sufficient for P because P_{12} simultaneously causes mental event M , which 'helps' bring about P .

Setting aside scruples about simultaneous causation, the real problem comes when we ask just how M helps bring about P . What the diagram shows is a contributing cause. In that regard M is no different from P_{11} . This may (or may not) make sense. But it is certainly not an example of the type of causation that Lowe claims to be offering. Recall that Lowe is proposing a model on which mental causes are indirect rather than direct. M is not supposed to intervene directly in the convergent causal tree. Rather it explains why there is a convergent causal tree at all — or, as he puts it in *Personal Agency* (2008, p. 36), why P is not a coincidence. But M cannot be doing any such job in the model depicted in Figure 1, since M is actually the causal product of one of P 's immediate physical causes.

This drives Lowe to the other horn of the dilemma, on which the mental act of choosing does not causally intervene in the branching tree of neural events. But now he needs to answer an even more difficult question: how can the mental act have any causal influence on the arm movement at all? *Ex hypothesi* the branching tree of neural events

is causally sufficient for the particular bodily movement that occurs (or at least, in a probabilistic universe, sufficiently fixes the chances of that movement). And yet Lowe is committed to it being the case that had the mental act not occurred, not only would that bodily movement not have occurred, but nor would there have been any bodily movement of that type at all. In virtue of what could this counterfactual dependence hold, if the mental act is causally insulated from the physical antecedents of the bodily movement? It certainly looks as though we have the mere statement of a counterfactual dependence without any account of what could possibly make it true.

In *Personal Agency*, Lowe does give a theological example to illustrate how his proposed account is at least possible on the assumptions that I am terming the second horn of the dilemma. Imagine a world in which it is true that physical events only have other physical events in their causal history and in which every physical event has a sufficient physical cause. Such a world would, he thinks, have no beginning in time. He continues:

And yet we could still ask of this world why *it* should exist or be actual rather than any other. One intelligible answer would be that this world was actual because God had chosen it to be actual. God's choice, then, would have caused it to be the case that a world containing certain physical facts was actual — and this would be mental causation of physical causal facts. (*ibid.*, p. 55)

This is a desperate move. Leaving aside the question of whether the theological scenario really is intelligible, taking this example literally would seem to require that ordinary mental causation operate outside space and time. If ever there was a cure that killed the patient this is it!

3. The Unity Argument

In Section 5.2 of *Personal Agency* Lowe offers what he terms the unity argument for NCSD. The argument is intended to show that the self cannot be identical to the body or to any part of the body (such as the brain or central nervous system, for example). Here is the argument:

- (1) I am the subject of all and only my own mental states.
- (2) Neither my body as a whole nor any part of it could be the subject of all and only my own mental states.
- (3) I am not identical with my body or any part of it.

Clearly the weight of the argument is being carried by premise (2).

In support of (2) Lowe observes that my body as a whole does need to exist in order for me to have all, or at any rate almost all, of the mental states I currently have. So, for example, if I were to lose the tip of the little finger on my left hand, that would have no implications for any of my mental states (except for any non-phantom sensations that I might feel at the tip of my little finger). Lowe then applies the following principle.

No entity can qualify as the *subject* of certain mental states if those mental states could exist in the absence of that object. (Lowe, 2008, p. 96)

Since he believes that the fingertip example shows that the overwhelming majority of my mental states could exist in the absence of my body, he concludes that my body cannot be the subject of those mental states. Since exactly the same argument could be applied to any part of my body (imagine, for example, my brain lacking a few brain cells), Lowe concludes that no part of the body could be the subject of all and only my own mental states.

The obvious objection to this line of argument is to deny that my body would cease to exist if I were to lose the tip of the little finger on my left hand. It seems much more plausible to say that my body would persist through such a change. After an involuntary digital amputation, for example, I would end up with the same body, slightly changed, rather than with a different body that almost completely overlaps with the old one. To hold the opposite, as Lowe does, would seem to entail that one's body disappears when a hair drops out or a toenail falls off.

Lowe acknowledges this objection and revises his defence of (2) to accommodate it. Let 'T' denote the set of occurrent thoughts that do not depend upon my body having the tip of the little finger of my left hand as a part, let 'O' denote my body minus the tip of the little finger of my left hand, and let 'B' denote my body as a whole. Now, Lowe asks, what reason could there be for identifying O rather than B as the subject of the thoughts in T? He finds none, claiming that the material difference between O and B is irrelevant to either of them being the subject of the thoughts in T. But then, he continues, it must either be the case that B and O are both the subjects of the thoughts in T or that neither of them is. Since thoughts cannot have two distinct subjects, and B and O are distinct objects, the first option is ruled out.

It would be dialectically unsatisfying simply to respond to this line of argument that only B is a genuine object. That would be to beg the

question against Lowe, who has written at great length on the metaphysics of identity and constitution. However, even if we put aside the question of whether O is a genuine object (as we probably should, since 'object' is not a term that has clear criteria of identity), it is not clear that on his own metaphysical principles Lowe is entitled to assume that O and B are equally eligible to be subjects of the thoughts in T. Lowe is firmly committed to the metaphysical significance of the category of substance. Although he does not, to my knowledge, directly consider the issue, I think that he is committed to holding that only substances can be the subjects of thoughts and experiences. This is the crucial difference between B and O. Only B is even a candidate to be a substance, because B is an appropriately organized and structured physical organism that falls under a sortal term with clear criteria of identity (the sortal 'human body').

It is certainly true that in a possible world W in which I had had the tip of the little finger on my left hand amputated, I would still count as a substance, because in that possible world my body would be appropriately organized and structured. Being a human body does not require possessing all one's fingertips. But it doesn't follow that in this world, where I do have all my fingertips, O, as a proper part of me, is a substance. In fact, it fails to qualify even by Lowe's own definition of substance, which is x is a substance if and only if it is not dependent for its existence upon anything else (Lowe, 1998, p. 138). O would not be dependent upon anything else in the counterfactual scenario, but in this world O is dependent for its existence upon the human body B of which it is a proper part.

For these reasons Lowe is not entitled to his crucial claim that it must either be the case that B and O are both the subjects of the thoughts in T or that neither of them is. And so the revised argument for (2) fails. The unity argument is ingenious but ultimately unsuccessful.

4. Individuating Psychological Modes: Lowe's Objection to Reductionism about the Self

Throughout his career Lowe complemented his positive arguments for NCSD with probing criticisms of what he took to be the principal alternatives. Since it is integral to his view of selves as simple substances that they be the subjects of psychological events, process, and states, a natural target for him was reductionism about the self. His

objections to reductionism clarify and sharpen his conception of what it is for a self to be a subject of experience.

The reductionist position was put most pithily by Hume in *A Treatise on Human Nature* when he described the self as ‘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’ (Hume, 1739–40/1978, p. 252). Hume continues:

The mind is a kind of theatre, where several perceptions successively make their appearance; pass, re-pass, glide away, and mingle in an infinite variety of postures and situations. There is properly no *simplicity* in it at one time, nor *identity* in different; whatever natural propensity we may have to imagine that simplicity and identity. (*ibid.*, p. 253)

The most celebrated modern exponent of reductionism (also known, for obvious reasons, as the bundle theory) is Derek Parfit (Parfit, 1984).³

Lowe uses the expression ‘psychological modes’ to cover mental events, process, and state. I will follow him in this convenient shorthand. Lowe’s basic objection to reductionism about the self is that psychological modes are necessarily ‘owned’. They cannot, he thinks, be identified and individuated in the way that the reductionist maintains. The reductionist needs psychological modes to be individuated impersonally — namely, without reference to the subject of which they are modes. But Lowe, like many others, thinks that this is impossible.

Such individual mental states are necessarily states *of persons*: They are necessarily ‘owned’ — necessarily have a *subject*. The necessity in question arises from the metaphysical-cum-logical truth that such individual mental states cannot even in principle be individuated and identified without reference to the subject of which they are states. (Lowe, 1996, p. 25)

Of course, no reductionist is going to accept that there really is such a metaphysical-cum-logical truth. So Lowe’s position ultimately rests upon his objections to impersonal criteria for individuating psychological modes.

³ I don’t actually believe that Parfit offers the only way, or even the best way, to formulate reductionism about the self. But it is the version that Lowe considers, and in the remainder of the paper ‘bundle theorist’ should be understood to refer to ‘Parfit-style bundle theorist’.

Perhaps the best-known proposal for individuating psychological modes comes from Donald Davidson, as an application of his general criterion for individuating all events. Davidson held that we can individuate mental events causally. Two events are identical iff they have the same causes and effects (Davidson, 1969). Plainly this causal conception of individuation can be applied within an impersonal conception of the world. So Lowe discusses Davidson's position at some length.

Before we look at the details of Lowe's discussion of Davidson we should recognize that Davidson himself abandoned this theory of events (Davidson, 1985). It is still worth considering, however, for two reasons. The first is that a recognizable descendant of the theory still lives on in the philosophy of mind in functionalist theories of mind, which typically identify mental states/events, including experiences, in terms of their typical causes and effects. The second, and for our purposes more important, is that Lowe's objection to Davidson is really a more general objection to a certain model of individuation (impredicative individuation) and it is highly plausible that any reductionist account of how to individuate mental events will be impredicative.

Lowe's principal objection is that the causal criterion of event identity is, if not strictly speaking circular, then at least impossible to apply — a charge that, as he freely admits, has been levelled many times at Davidson (for a pithy example see Quine, 1985). Since the relata of the causal relation are themselves events, in order to apply the criterion to individuate an event we would have already had to have individuated the events that are its causes and effects. Davidson's definition of event identity is an example of impredicativity, whereby something is defined with reference to a totality of which it is an element. Philosophers are divided on whether impredicative definitions are acceptable as definitions, or whether they are viciously circular (this is a particularly lively topic in the philosophy of mathematics). But Lowe agrees with Quine that impredicative individuation is a non-starter.⁴

⁴ Actually Lowe is not entirely on the same page as Quine with respect to impredicative individuation. He discusses Quine's views in Lowe (1989). What Quine should have said, according to Lowe, is that impredicative individuation is impermissible in the absence of a supporting theory that will allow an impredicative principle to be applied (in the way that the impredicative axiom of extensionality in Zermelo-Frankel set theory

This proscription on impredicative individuation needs closer examination, however. The topic of individuation is one that Lowe has explored at length throughout his career and his own discussion is an excellent starting point. We can start with some general points about individuation that would be accepted by most philosophers and certainly by Lowe. First, individuation is relative to a given category or sortal. We need to know the kind of thing that we are dealing with before we can specify how that thing is individuated. Second, there is a close connection between individuation and criteria of identity. To individuate an object is to give criteria that determine whether an object is the same at different times and that will settle the question of whether at a given time we have two objects of a given kind or just one.

In *More Kinds of Being* Lowe gives the following as the general form for criteria of identity.

- (1) If x and y are ϕ s, then x is identical to y if and only if x and y satisfy $C(\phi)$ (Lowe, 2009, p. 16).

This general schema can be read both at a time and over time, and it is sortal-relative — that is to say, it gives a criterion for identity for x and y relative to some sort or category ϕ under which they both fall.

The general form (1) allows us to clarify what would count as impredicative individuation. We have impredicativity when the condition $C(\phi)$ makes reference to a set or totality in which x and y both feature. Davidson's criterion for event identity certainly counts as impredicative in this sense. The condition he proposes where ϕ is the category of events is that $x = y$ iff x and y have the same causes and effects. The most natural way of formulating that condition would be

- (2) $x = y$ iff $\forall z [(z \text{ causes } x \leftrightarrow z \text{ causes } y) \ \& \ (x \text{ causes } z \leftrightarrow y \text{ causes } z)]$.

In (2) we assume that the universal quantifier \forall ranges over all events. Since x and y are events, they fall within the range of \forall and so (2) is impredicative.

Why is this supposed to be a problem? Here is Lowe.

Briefly, the trouble with Davidson's criterion is that if (as Davidson himself proposes) the causes and effects of events are themselves events

is supported by the remaining axioms). These differences are not relevant to the current discussion, however.

then the question of whether events e_1 and e_2 have the *same* causes and effects (and hence turn out to be the same event according to the criterion) is itself a question concerning the identity of events, so that in the absence of an independent criterion of event identity Davidson's criterion leaves every question of event identity unsettled. (Lowe, 1996, pp. 27–8)

I find this reasoning unconvincing. Consider the following reformulation in which talk of questions is replaced by talk of facts.

Briefly, the trouble with Davidson's criterion is that if (as Davidson himself proposes) the causes and effects of events are themselves events then the *fact* of whether events e_1 and e_2 have the *same* causes and effects (and hence turn out to be the same event according to the criterion) is itself a *fact* concerning the identity of events, so that in the absence of an independent criterion of event identity Davidson's criterion leaves every *fact* about event identity unsettled.

This, I submit, would not be compelling. There is nothing mysterious about the idea that a fact about the identity of an object of a given type depends upon facts about the identity of other objects of the same type. Consider for example the well-known theory of origin essentialism proposed by Saul Kripke. According to Kripke, a human being's origins are essential to that human being (Kripke, 1980). One way of formulating this would be to say that it is essential to any human being that they should have developed from the actual zygote that they did develop from. Since a zygote is the immediate product of a sperm cell originating (essentially) in a man fertilizing an egg cell originating (essentially) in a woman, it follows that it is essential to any human being that they have the parents that they actually have. This has obvious implications for how we think about identity. It means that x is the same human being as y iff x and y have the same parents. But parents are human beings, and so we have a fact about the identity of a human being dependent upon facts about the identity of human beings. Such dependence would be unacceptable by the reformulated argument. But there's absolutely nothing wrong with it. Whatever problems there might be with Kripke's origin essentialism this is not one of them.

The lesson to draw from this, I think, is that we need to distinguish two different senses of the term 'individuation'. On the one hand, individuation can be taken metaphysically. From a metaphysical point of view individuating an object is a matter of specifying what makes that object the object that it is — giving the criteria that determine that object's identity at a time and over time. On the other we can think

about individuation as an epistemic undertaking, as a process by which we actually go about establishing whether there really is one object when there appear to be two, or whether an object really has persisted over time. The contrast is between what makes it the case that $x = y$ or $x \neq y$ and how we go about determining whether $x = y$ or $x \neq y$.

The objection to impredicative individuation that we are discussing depends upon framing individuation epistemically. His objection is really that it would be in practice impossible to apply Davidson's criterion for event identity — in order to apply it to settle a question about the identity of two events we would already have to have settled questions about the identity of the events that are the causes and effects of the given event(s). But the objection misses its target. The issue we started with was a metaphysical issue, not an epistemic one. The reductionist is trying to give the identity conditions of mental events without reference to selves or persons. That requires explaining what makes mental events the same or different, not how we go about establishing whether they are the same or different. And, as the example of Kripke's origin essentialism shows, from a metaphysical perspective impredicative individuation can be perfectly acceptable.

Lowe might object that metaphysical and epistemic individuation cannot be so easily separated. He could argue, for example, that a metaphysical identity criterion must be something that we are in principle able to apply, and so to the extent that impredicativity is an epistemic problem it is equally a metaphysical one.

However, there are different contexts in which we might be thinking about individuation. In some situations we have independent modes of access to the object under consideration. The origin essentialism case is like this. We have many different ways of identifying human beings, and we can use them to apply the essentialist criterion of identity. We do not need to investigate the parents of x and y 's parents in order to determine whether or not x and y have the same parents. The same holds, I submit, for mental events. We have sufficient grip on what mental events are in order to be able to work out for any mental events w and z whether or not they have the same causes and effects without getting into an endless regress of identifying and comparing the causes and effects of the causes and effects, and so on, of w and z . In situations such as these we do not need to apply criteria of identity all the way down, because we can reply on a more intuitive grasp of, say, mental events. From an epistemic point of view, we use a formal principle of individuation such as Davidson's in order to help

resolve problem cases of mental event identity. We do this, though, in the context of being able to resolve ordinary, non-problem cases, and that ability allows us to apply impredicative criteria of identity.

There are situations, though, where we do not have an independent mode of access to the objects under consideration. This may well be the case in mathematics. If a class of mathematical objects is introduced impredicatively and we have no independent grasp of how to individuate objects in that class, then there seems to be room for doubts about how we might (epistemically) individuate objects in that class. The issue is very much alive in the philosophy of mathematics (see Shapiro, 2000, for discussion and further references). But we are obviously in a very different epistemic position with respect to mental events than we are with respect to mathematical objects.

It is true, as a referee observed, that in other publications (e.g. Lowe, 2002, pp. 226–8, and 2010) Lowe offers a different type of objection to Davidson's causal criterion for event identity. He describes a simple possible world with a handful of events where the causal relations are in effect symmetrical, so that (he argues) the causal criterion fails uniquely to individuate the events in that world. This is certainly a metaphysical argument, not an epistemic one. But, as emphasized earlier, the important point is not really whether Davidson's theory of identity can ultimately be maintained. I have been focusing on the legitimacy of impredicative individuation, because it is highly plausible that any reductionist proposal for individuating mental events will have to be impredicative. And what I have tried to show is not that Davidson's theory should be maintained, but rather that Lowe's objection to impredicative individuation (illustrated by Davidson, but not exclusive to Davidson) rests upon an illegitimate appeal to epistemic criteria.

Let me take stock. His attack on reductionism about the self is an important plank in Lowe's case for NCS. In this section we have been reviewing the objection to reductionism that he levels in *Subjects of Experience* (1996). Lowe argues that reductionism is false because there is no possibility of individuating mental events without reference to selves as their owners. He considers a causal criterion derived from Davidson's well-known discussion of event identity and objects, in essence, that such a criterion would be illegitimately impredicative. Against this I have argued that impredicative individuation is perfectly acceptable for mental events. This falls short, of course, either of a definitive formulation of reductionism or of a convincing argument in favour of reductionism. Nonetheless, reductionism about the self

remains in play and we still lack compelling support for NCSD. In the final two sections I turn to another of Lowe's anti-reductionist arguments.

5. Self-Reference and the Boundaries of the Self

In his paper 'Self, Reference, and Self-Reference' (Lowe, 1993), a precursor of the final chapter of *Subjects of Experience*, Lowe offers an account of selfhood that ties it to the self's *de re* knowledge of its own thoughts and experiences. It is, he argues, constitutive of the concept of selfhood that a self be able to make direct demonstrative reference to his or her own occurrent thoughts and experiences, as well as to those parts of the body that can be moved at will. In fact, those parts of the body that can be moved at will are the only physical objects to which Lowe thinks it possible to make direct demonstrative reference, so that the semantic limits of direct demonstrative reference (DDR) track the metaphysical boundary between self and world. He then goes on to develop a further argument against reductionism about the self. This section discusses his conception of direct demonstrative reference. The final section evaluates the anti-reductionist argument.

Standard accounts of reference make a contrast between DDR and indirect reference according to whether the object referred to is sensibly present. We can make DDR both to our own thoughts and experiences and to appropriately situated objects in the external world. According to Lowe, however, reference to non-bodily physical objects is indirect, where this means that a particular demonstrative reference is fixed (implicitly or explicitly) by one or more independent acts of reference involving definite descriptions. Only when we refer to what we are presently thinking or experiencing, or to those parts of our body that we can move 'at will', can we refer demonstratively without relying upon such independent acts of reference.

The standard view permits DDR to non-bodily physical objects such as chairs, provided they can be sensibly discriminated. But, Lowe asks (1993, pp. 28–30), how do I know to which chair I am referring? Only, he replies, through some definite description, such as 'the chair on which I am now fixing my gaze' or 'the chair which I can see on my extreme left out of the corner of my eye'. Such definite descriptions involve an implicit reference to myself, and that is enough to make the reference to a chair indirect. In contrast, no such definite descriptions are required when referring to my present conscious thoughts and experiences.

In the case of a chair Lowe thinks that a definite description is required to provide completely unambiguous reference-fixing. The contrast he draws rests on the claim that reference to one's mental states is always unproblematic and unambiguous. The need for disambiguation in reference to chairs arises because of the possibility of confusion, because there may be more than one chair in front of one (*ibid.*, pp. 28–9). But a similar possibility of confusion exists in reference to one's mental states. Perhaps my toothache is playing up just as my sprained ankle twinges, or perhaps I've just incurred multiple fractures. In these and many other cases one has a plethora of distinct pains, not one composite pain. And, for Lowe's purposes, there is surely no relevant difference between such a situation and one in which one is confronted by a plethora of chairs. Disambiguation is clearly required for reference-fixing in both cases.

How might such disambiguation be provided for multiple pains? Well, I might be referring to the pain to which I am attending. But here we seem drawn to a definite description, 'the pain I am attending to', and with it an indirect reference to myself. Alternatively, I might be referring to a pain in a particular location, 'the pain that I can feel in my arm'. This too brings with it an independent demonstrative reference to myself, as well as to my arm. Or perhaps reference could be fixed through qualitative feel ('the most excruciating pain', for example). But this would be equally unsuccessful, because an independent demonstrative reference is needed to fix the range of the quantifier. I am not referring to the most excruciating pain in the world or in the room, after all, but just to the most excruciating pain that I am feeling now. To appreciate this, consider how I might disambiguate the statement 'that pain is excruciating', uttered when I am lying in a hospital bed looking at the fellow victims of a multiple car crash.

Of course, these examples yield what might seem a trivial sense in which reference to my pains counts as indirect. But it is in precisely this sense that reference to perceptually presented objects is indirect, according to Lowe. If he thinks that I might need to make clear which chair I am referring to by means of a definite description like 'the chair on my right' or 'the chair in front of me', then he must accept that I might need to make clear which pain I am referring to by means of a definite description like 'the sharp pain I am feeling now' or 'the pain in my right hand'. And if he thinks that the former count as indirect reference, then so too must the latter.

Lowe cannot object that a theory of reference should not be based on extrapolation from these unusual cases of multiplicity and potentially ambiguous demonstrative reference, for he explicitly holds the opposite: ‘Cases of actual observable multiplicity only serve to make the indirect nature of all such demonstrative reference more evident’ (*ibid.*, p. 30). In the case of perceptually based demonstrative reference to chairs Lowe thinks it legitimate to conclude that such reference must always be indirect from occasional cases of multiplicity. Hence he cannot avoid doing so in the case of reference to one’s own thoughts and experiences.

Lowe does on one occasion wonder whether the possibility of multiplicity might not also exist in the case of conscious mental states, but dismisses it on the grounds that one cannot have qualitatively identical but distinct thoughts and experiences (*ibid.*, p. 30 — see also Lowe, 2008, pp. 26–31). I am not convinced by this. Nobody would deny that I can, at two different times, have two qualitatively identical pains, one in my right foot and one in my left foot. So why should it be impossible for me to have those two qualitatively identical pains at the same time? There seems no *a priori* reason for demanding that they suddenly merge into one pain, or acquire a qualitative difference. The view that all pains must be qualitatively different is surely wildly implausible. So why should the view that all the pains one is experiencing at any given moment must be qualitatively different be any less implausible? But in any case, as I have shown, problems of multiplicity arise even when qualitative identity is not assumed, because we saw that an implicit self-reference can be required even when a pain is being identified qualitatively (as ‘the excruciating pain that I am feeling now’).

Lowe’s paper suggests how he might respond. Considering the possibility that one might identify a pain via the definite description ‘the toothache that I am now feeling’, he asks ‘...can it seriously be suggested that demonstrative reference by me to my own current toothache necessarily relies, even if only implicitly, upon an independent act of reference to myself? That would seem to suggest that without implicitly specifying to myself that it is my toothache to which I intend to refer as “this pain”, I might mistakenly be referring to someone else’s...’ (Lowe, 1993, p. 21). He finds this completely absurd (because he holds it a logical truth that one can only feel one’s own pain). But even granting that there is no ambiguity about who the pain belongs to, there remains an ambiguity about which one it is of the pains that (unquestionably) belong to me. This ambiguity about

which pain it is precisely matches the ambiguity Lowe thinks there is when I have several chairs in front of me. So, if an implicit self-reference is required in the latter case, it must also be required in the former.

He also suggests that no reference to oneself is involved in demonstrative reference to one's body (so that one could, despite my suggestion earlier, directly refer to the pain in one's arm, and hence individuate pains by their bodily location): '...it seems absurd to suppose that I need to specify that a hand is mine in order to make it an object of my volitions — for the idea that I might move another's hand at will (that is, as a basic action) is just incoherent' (*ibid.*, p. 32). But ruling out the possibility that it actually belongs to somebody else is not the only reason one might have for appealing to such descriptions to fix the reference of 'this hand'. For example, one might think that in some situations the reference of 'this hand' (e.g. an answer to the question 'which hand?') can only be fixed by a definite description such as 'not the hand I write with'. I might be confused about which hand I am going to pick up the potentially scalding object with until I think of it under the description 'not the hand I write with'. Here, as in the case of multiple pains, the ambiguity is between my two hands, rather than between my hand and somebody else's hand.

Lowe's distinction between direct self-reference and indirect reference to non-bodily physical objects cannot be maintained on the grounds that he presents for it. His construal of direct reference has the consequence that not even reference to our own thoughts and experiences count as instances of DDR. So, assuming that a distinction between indirect reference and DDR is a necessary feature of a satisfactory theory of reference, the standard theory is still the only serious candidate.

6. Direct Self-Reference and Reductionism

Lowe uses his theory of demonstrative reference against the reductionist view of the self, objecting that reductionism about the self makes first-person self-reference into a form of indirect reference. Section 5 shows that, if this is a genuine criticism of reductionism, it is an equally valid criticism of Lowe's own theory of DDR. But his anti-reductionist argument can be freed from his theory of DDR, and is independently interesting.

The argument runs as follows (Lowe, 1993, pp. 25–7). It is a defining feature of selfhood that the self knows necessarily that it is the unique subject of its present conscious thoughts and experiences. This knowledge, expressible in propositions such as ‘*This pain is my pain*’, is genuine and informative. But on the bundle theory such genuine and informative self-knowledge would not be available, because the bundle theorist can only refer to the self indirectly, through the thoughts and experiences to which it reduces. If the self can only refer to itself indirectly, by means of independent reference to these experiences and thoughts, then ‘*This pain is my pain*’ will come out as trivially true, because it is trivial that *this* pain is one of *these* pains.

An initial response might be that the bundle theorist *can* provide an account of indirect self-reference which will make propositions such as ‘*This pain is my pain*’ into informative self-knowledge, by construing ‘*This pain is my pain*’ as ‘*This pain is co-conscious with other thoughts and experiences to which DDR could now be made*’. This is not trivial in the way that ‘*This pain is one of these pains*’ is trivial, because *this* pain is not included in the set of other thoughts and experiences with which it is co-conscious. And nor should it be thought that ‘other thoughts and experiences...’ is effectively equivalent to ‘*these* other thoughts and experiences...’, so that the original difficulty reappears. It is trivially true that a pain is co-conscious with itself, because co-consciousness is a reflexive relation, but this is not what is doing the work here. The point of the bundle theorist’s proposal is precisely that *this* pain is co-conscious with a range of mental states that are distinct from itself and to which DDR *could* now be made.

But this is not quite right yet, since at any given moment any thought or experience is such that DDR could now be made to it. We clearly do not want the result that all thoughts and experiences are co-conscious with each other. Nor, on the other hand, can the bundle theorist get round the problem by reformulating ‘*This pain is my pain*’ as ‘*This pain is co-conscious with other thoughts and experiences to which I could now make DDR*’. That would be to concede the point to the anti-reductionist, by making an appeal to the self ineliminable.

But Lowe’s bundle theorist does have a way out. There is more to thoughts and experiences being co-conscious than the possibility of making DDR to them. For example, the co-consciousness of mental states is manifest in various dispositions to draw conclusions — if a desire for *x* is co-conscious with a memory that *x* is to be found at *p*,

then there will be a disposition to infer that *p* is a desirable place to go to. By the same token, and more generally, the co-consciousness of mental states brings with it the *a priori* ability to conjoin them in a more inclusive mental state. Take the judgment 'I am in pain', or, as a reductionist might say, 'there is pain here' — I won't spell out the impersonal rewording in the following. Suppose it is co-conscious with the judgment 'I am tired', then without recourse to experience the conjunctive thought 'I am in pain and I am tired' can be formulated. This is obviously not the case if you think 'I am in pain' and I think 'I am tired' (taking this as a paradigm case of two thoughts that are not co-conscious). The bundle theorist can take this capacity to form *a priori* conjunctive thoughts as constitutive of co-consciousness, so that two thoughts are co-conscious iff thinking of them makes it possible to form *a priori* a thought conjoining them.

Understanding co-consciousness in this way enables the bundle theorist to construe 'This pain is my pain' as informative self-knowledge. He can reformulate it as 'This pain is a member of a set of thoughts and experiences which supports *a priori* conjunctive thoughts'. Because this is neither trivial nor uninformative, Lowe's argument from self-reference against reductionist accounts of the self cannot be sustained.

Lowe's non-Cartesian substance dualism is a provocative and ingeniously defended theory. I remain unconvinced both by his positive arguments for NCS and by his criticisms of reductionism. However, he has done us all a service by presenting his theory so clearly and by forcing defenders of opposing views to clarify their own thoughts and arguments.

Acknowledgments

The special issue in which this paper is published had its origins in a conference in honour of E.J. Lowe at Durham that sadly turned into a memorial conference. I was unable to attend the conference due to prior commitments and I am very grateful to Mihretu Guta and Sophie Gibb for allowing me to contribute a paper to the special issue. The paper was much improved by the acute comments of two referees.

References

- Adams, M.M. (1982) Universals in the early fourteenth century, in Kretzmann, N., Kenny, A. & Pinborg, J. (eds.) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press.
- Davidson, D. (1969) The individuation of events, in Rescher, N. (ed.) *Essays in Honor of Carl G. Hempel*, pp. 216–34, Dordrecht: Reidel.
- Davidson, D. (1985) Reply to Quine on events, in LePore, E. & McLaughlin, B.P. (eds.) *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, pp. 120–3, Oxford: Basil Blackwell.
- Hume, D. (1739–40/1978) *A Treatise of Human Nature*, Oxford: Oxford University Press.
- Kripke, S. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press.
- Lowe, E.J. (1989) Impredicative identity criteria and Davidson's criterion of event identity, *Analysis*, **49** (4), pp. 178–181.
- Lowe, E.J. (1993) Self, reference, and self-reference, *Philosophy*, **68**, pp. 15–33.
- Lowe, E.J. (1996) *Subjects of Experience*, Cambridge: Cambridge University Press.
- Lowe, E.J. (1998) *The Possibility of Metaphysics: Substance, Identity, Time*, Oxford: Oxford University Press.
- Lowe, E.J. (2002) *A Survey of Metaphysics*, Oxford: Oxford University Press.
- Lowe, E.J. (2008) *Personal Agency: The Metaphysics of Mind and Action*, Oxford: Oxford University Press.
- Lowe, E.J. (2009) *More Kinds of Being: A Further Study of Individuation, Identity, and the Logic of Sortal Terms*, Oxford: Wiley-Blackwell.
- Lowe, E.J. (2010) On the individuation of powers, in Mormodoro, A. (ed.) *The Metaphysics of Powers: Their Grounding and their Manifestations*, London: Routledge.
- Parfit, D. (1984) *Reasons and Persons*, Oxford: Oxford University Press.
- Quine, W.V.O. (1985) Events and reification, in Lepore, E. & McLaughlin, B. (eds.) *Actions and Events: Perspectives on the Philosophy of Davidson*, Oxford: Blackwell.
- Shapiro, S. (2000) *Thinking about Mathematics: The Philosophy of Mathematics*, Oxford: Oxford University Press.