Leibniz, Locke and I

SIMON BECK

1. Introduction

In his *New Essays on Human Understanding*, Leibniz presents a sharp attack on Locke's theory of personal identity. Matching Locke's thought-experiments with those of his own, Leibniz seeks to show that our identity cannot rest on matters of consciousness alone—being the same person is rather a matter of the continued existence of an immaterial substance.

Nowadays the consensus seems to be that Leibniz loses this particular battle with Locke, although that is more because of the unpopularity of immaterial substances than the weakness of Leibniz's argument. Nevertheless, I wish to draw attention to some contemporary thinkers who—while eschewing the immaterial substances—are sympathetic to the kind of argument Leibniz offered. This leads to a dilemma: on the face of things, we cannot be sympathetic to both Leibniz and Locke.

I will investigate a theory which would resolve this dilemma—giving up something which each held dear, but not everything. The resultant theory leans unpopularly more towards the side of Leibniz than Locke but can point to the authority of no less than David Wiggins. In the end, though, I wish to argue that there is another solution to the dilemma; a solution which claims that Leibniz and Locke are arguing at cross-purposes, and that their arguments are best seen as answers to different questions. This solution still requires the giving-up of widely-held beliefs, but involves significantly less cost than the alternatives.

2. Locke's case

In the second edition of his *Essay*, Locke argues that there is a distinction to be drawn between humans and persons, and that the identity conditions of the two categories are different. He writes:

> should the soul of a prince, carrying with it the consciousness of the prince's past life, enter and inform the body of a cobbler, as soon as deserted by his own soul, every one sees he would be the same person with the prince, accountable only for the prince's actions: but who would say it was the same man? (Locke, 1694: II, pp. xxvii, 15)[1]

Considering this example leads him to the following conclusion as to what constitutes the identity of a person:

> as far as any intelligent being can repeat the idea of any past action with the same consciousness it had of it at first, and with the same consciousness it has of any present action; so far it is the same personal self. (Locke, 1694: II, pp. xvii, 10)

Locke realises that there is a gap in this argument. The thought-experiment militates only against bodily or other physical continuity as being necessary for personal identity; but this does not directly support the kind of memory criterion he espouses. In the case of the prince and the cobbler, as he outlines it, it is quite clear that the work of swapping 'consciousnesses' between the two bodies is done by the *souls* of the individuals involved. What then prevents the identity of a person consisting in the continued existence of an immaterial substance—a soul as envisaged by Descartes?

Locke's argument against the idea of the self being an immaterial substance is to argue that the self is not a *substance* at all, immaterial or otherwise. In this, he makes use of another thought-experiment: he holds that were some-

one to insist that his soul were that of Socrates—that is, he and Socrates shared their immaterial substance—while being unable to remember any of Socrates's actions or thoughts, we would deny that he and Socrates were the same person. And we would still do so even if it could somehow be proved that there was an immaterial substance shared between them (1694: II, pp. 27, 14).

This view, that a person's identity relies upon connections of memory and not upon the underlying existence of any sort of substance, is the view that Leibniz takes exception to in his *New Essays*.

3. Leibniz's reply

Leibniz's account of personal identity is fairly complex, and shares a number of points with Locke's (Curley, 1982; Jolley, 1984; Scheffler, 1976).[2] Nevertheless, Leibniz denies that connections of memory are necessary for identity and insists on the importance of immaterial substances.

First, with regard to memory, Leibniz offers a *reductio ad absurdum*.

> Must it not be agreed that after some passage of time or some great change one may suffer a total failure of memory? They say that Sleidan before his death forgot everything he knew, and that there are plenty of other examples of this sad phenomenon. Now, suppose that such a man were made young again, and learned everything anew—would that make him a different man? So it is not memory that makes the very same man. (Leibniz, 1765: p. 114)

His case for Locke's not taking seriously enough the role played in identity by immaterial substance involves a more complex and much more lively thought-experiment. The voice is that of Leibniz's representative Theophilus attacking Philalethes and his Lockean 'authorities':

> Here is something we could much more fittingly suppose: in another region of the universe or at some other time there may be a sphere in no way sensibly different from this sphere of earth on which we live, and inhabited by men

each of whom differs sensibly in no way from his counterpart among us. Thus at one time there will be more than a hundred million pairs of similar persons, i.e. pairs of persons with the same appearances and states of consciousness. God could transfer the minds, by themselves or with their bodies, from one sphere to the other without their being aware of it; but whether they are transferred or left where they are, what would your authorities say about their persons or 'selves'? Given that the states of consciousness and the inner and outer appearances of the men on these two spheres cannot yield a distinction between them, are they two persons or are they one and the same? . . . Since according to your theories consciousness alone distinguishes persons, with no need for us to be concerned about the real identity or diversity of substance . . . what is to prevent us from saying that these two persons who are at the same time in these two similar but inexpressibly distant spheres, are one and the same person? Yet that would be a manifest absurdity. (Leibniz, 1765: p. 245)

Real identity is not a matter of being able to remember experiences, then. While Leibniz has his differences with Descartes, he is forced to agree that identity is ultimately a matter of being a distinct immaterial substance or soul.

4. The exchange considered

Leibniz's first argument is not a particularly strong one, and more recent discussions of personal identity point to how the objection can easily be avoided. In the first place, he simply takes it as obvious that Sleidan survived amnesia without giving us anything like a convincing reason as to why we should concur with this judgement. Nevertheless, there are cases when we are tempted to agree that a person retains their identity despite amnesia. But this temptation only occurs when other of their psychological characteristics remain—when, for instance, they retain characteristic beliefs, projects, sense of humour and so on. What this suggests is that it is a mistake to put all the weight that Locke does

on memory; personal identity must be held rather to depend on psychological continuity in a much broader sense—memory being but one (albeit an important one) of the factors in the required continuity.

Leibniz does have more to say when he puts this first objection, however. He also presents an attack on the thought-experiment on which Locke's case rests.

> as for the fiction about a soul which animates different bodies, turn about, with the things that happen to it in one body being of no concern to it in the other: that is one of those fictions which go against the nature of things—like space without body, and body without motion—which arise from the incomplete notions of philosophers, and which vanish when one goes a little deeper. (Leibniz, 1765: p. 114)

This has the form of an interesting objection—for it is indeed a grave methodological mistake to gloss over deep impossibilities by using too much fictional licence to evince intuitive agreement regarding a thought experiment.[3] But for the objection to work, the onus is on Leibniz to explain what deep impossibilities Locke is hiding from us in his story of the prince and cobbler swapping bodies. What he says is this:

> it must be borne in mind that each soul retains all its previous impressions, and could not be separated into two halves in the manner you have described: within each substance there is a perfect bond between the future and the past, which is what creates the identity of the individual. (1765: p. 114)

But this amounts to no more than Leibniz baldly stating his opposing view; it is not an *argument* as to why souls or psychologies cannot swap bodies. As a result, Leibniz's first argument presents no compelling case against a Lockean view of identity.

His second argument is not so easily dismissed, however. In describing his twin-earth, Leibniz does present a substantive argument. Nor is this simply an attack on memory alone, to be warded off by appeal to other psychological features. For here, Leibniz presents individuals who are exactly alike in *all* physical and psychological aspects; they cannot be distinguished by reference to memory,

but neither can they be by reference to any other psychological facts. Leibniz's central charge seems to be that, following Locke's criterion of personal identity, we are bound to say that a person and her counterpart on twin-earth are one and the same person since their psychologies are identical. There are manifestly two people here, but no psychological difference in which such a distinction can be grounded.

On the face of things, then, we are faced with something of a conundrum. Locke's prince and cobbler thought-experiment has a great deal of intuitive appeal, and Leibniz's criticism of it does nothing to undermine this. Yet Leibniz's twin-earth case evokes intuitions which seem to be strongly at odds with those to which Locke appeals.

I wish to argue below that we do indeed face a conundrum along these lines, and to offer a solution which sympathizes with aspects of both Locke's and Leibniz's accounts. Nevertheless I must admit that, despite its ingenuity, Leibniz's twin-earth thought-experiment does not offer argument enough to satisfy us that there is really a problem here. For while it is true that there is no qualitative psychological difference between A on earth and her counterpart A' on twin-earth, it is not at all clear that this forces the Lockean to confess that he has defined A and A' as one person and to thus admit the need for immaterial substances as the carriers of identity. At the worst, what the Lockean has to do is to admit that sometimes psychological features alone are insufficient for identity. Less damaging would be to argue that the distinct bodies of A and A' mark their psychologies as numerically distinct. At the worst, then, this would be a step back from Locke's initial position, but even then not a step into Leibniz's camp.

5. Reductionism and nonreductionism

The exchange between Locke and Leibniz is an early instance of the debate which still continues between reductionist and nonreductionist views on personal identity. Reductionists, as did Locke, argue that personal identity can be reduced to more familiar relations—relations of physical or (more usually) psychological continuity—and that the world can be completely described without reference to

persons or their identity. Nonreductionists take the Leibnizian line in holding that attempts at such a reduction fail; rather, personal identity is a relation *sui generis*, and one which is simple. Any attempt to reduce it to some more fundamental relation will lead either to contradiction or will leave out something important. Nonreductionists may or may not agree that a person is an immaterial substance like a Leibnizian monad, but at least in this much they agree with Leibniz.

So far, the discussion of the Locke–Leibniz exchange between reductionist and nonreductionist has followed the popular line in arguing that the nonreductionist case is not particularly convincing. But the fault may just lie in the particular example Leibniz chose— for there are contemporary arguments not far removed from Leibniz's own which might well give us cause to think again.

One comes from Geoffrey Madell. He suggests that there are straightforward and familiar considerations which give strong support to the notion that personal identity is something irreducible.

> the following two thought-experiments are equally intelligible: (a) that I might not have existed, but someone having *exactly* the life that I have had might have existed instead; (b) I might have had a *totally* different life, even to the extent of being born centuries earlier. (Madell, 1981: p. 79)

As with Leibniz's example, there are no psychological facts which can explain the apparent identity between myself as I am now and the individual with a totally different life. Given the intuition that I could have been different in this way then, Madell contends, my identity is a matter over and above any psychological facts.

Brian Garrett presents another argument in a Leibnizian vein. Making use of Nathan Salmon's 'four worlds paradox' (Salmon, 1982: appendix 1, sec. 28), he argues that there can be differences of identity which are not grounded in any physical or psychological similarities, differences or continuities. And that seems to mean that personal identity is ultimately a relation which is not reducible to any such physical or psychological facts. Here is a version of Garrett's argument (Garrett, 1991: pp. 368–9).
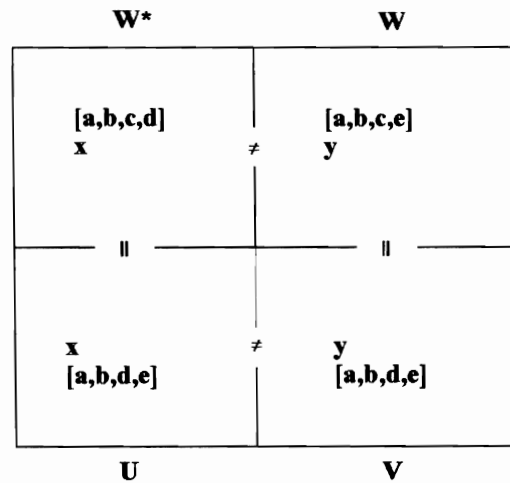
Assume, for ease of exposition, and since nothing crucial to the argument depends on it, that you were originally made up of four basic parts. Thus there is an individual, x, in the actual world, w^*, made up of parts a,b,c and d. Assume that there is another world, w, in which an individual y, who is not identical to x, exists, made up of parts a,b,c and e. This is not controversial: someone else could have existed who was originally made up of matter very similar, but not identical, to your own make-up. It is also hardly controversial to assume as well that you might have been made up of slightly different matter. That means that we can say that there is a world, u, in which x exists but is made up of parts a,b,d and e.

However, according to the uncontroversial assumption just invoked, there is also a world, v, in which y exists made up of a,b,d and e. Now, as a result of the transitivity and necessity of identity, y in v is not identical to x in u, since y in v is identical to y in w, while x in u is identical to x in w^*, and x in w^* is *not* identical to y in w. What all this means is that x in u and y in v are distinct individuals *even though there is nothing physical or psychological in which this non-identity can be grounded.* And thus identity cannot ultimately be a physical or psychological matter.

6. One attempt to solve the dilemma

The arguments of Madell and (especially) Garrett bring us back to the dilemma which appeared initially between the consequences of Locke's thought-experiment and those of

Leibniz's twin-earth case. Madell and Garrett present apparently compelling cases to the end that personal identity cannot be reduced to psychological continuity, whereas Locke presents an apparently compelling case that it can. Obviously we cannot have it both ways. Garrett has suggested that there is a solution to this dilemma, and that solution will be the focus of this section.

Garrett's way out is to acknowledge that identity cannot be *reduced* to psychological or physical continuity. Nevertheless, what Locke's case shows is that it can be *analysed* in those terms. The trick which is to work the solution is a distinction between reduction and analysis based on the realization that reducing a concept to others is not the only way of analysing it. Here Garrett can appeal to the authority of David Wiggins:

> No reduction of (sameness or identity) has ever succeeded . . . Nor is it called for, once we realize how much can be achieved in philosophy by means of elucidations which *use* a concept without attempting to reduce it, and, in using the concept, exhibit the connexions of the concept with other concepts that are established, genuine collateral and independently intelligible. (Wiggins, 1980: p. 4)

What Locke would be held to establish, then, is that the concept of personal identity—though not reducible to the concept of memory—has special connections to that concept, a concept which is indeed independently intelligible. This may not be as much as Locke had hoped for, but it is nevertheless philosophically valuable. And while Leibniz may well have been wrong about one's identity depending on the immaterial substance of which he thought one is composed, he was correct that identity is not reducible to consciousness. The clash between thought-experiments leads us to a sort of *analytic nonreductionism.*

7. Why Garrett's solution is problematic

Garret's solution is attractive, and the point that analysis can be interesting even when it is not reductive is well taken. All the same, the general wariness with which nonreductionism has been treated must be a warning for us to be careful here. I suspect that nonreductionism's bad name stems mainly from its traditional links with dualism, and it is important to note that neither Madell nor Garrett are committed to immaterial substances by their arguments. But even with this reassurance, it is still worth asking whether the case for nonreductionism has been made.

Both Madell and Garrett conclude, with Leibniz, that our identity over time cannot be reduced to a form of psychological continuity. I argued that Leibniz failed to make this point stick and, although their cases are not so easily faulted as Leibniz's, I suggest that they are ultimately no more successful than he was on this point.

The reductionist view is that X (at time t) = Y (at t-n) iff X is uniquely psychologically continuous with Y. A reductionist like Parfit would fill in the details along the following sort of lines.

> *Psychological continuity* is the holding of overlapping chains of *strong* connectedness . . . For X and Y to be the same person, there must be over every day *enough* direct psychological connections. Since connectedness is a matter of degree, we cannot plausibly define precisely what counts as enough. But we can claim that there is enough connectedness if the number of connections, over any day, is *at least half* the number of direct connections that hold, over every day, in the lives of nearly every actual person. (Parfit, 1984: p. 206)

Leibniz's *reductio* of this view was dissolved by pointing out that we *could* distinguish A' on earth from A on twin-earth using only the resources available to reductionists like Locke and Parfit. Madell's appeal to different *worlds* as opposed to different *planets* might seem to block a similar response to his case for non-reductionism. But how does the possibility of my having had a different life imply that psychological continuity is insufficient as a criterion for identity over time? The reductionist can explain that what makes it *my* life now at t as it was then at t-n is the psychological continuity outlined. That the particular psychological connections which make up this continuity could have been different is neither

here nor there; just as the fact that the old oak might have been made of different matter does not prevent it being the same tree as the young sapling which stood there 100 years ago. And nobody would suggest that because of this fact about oak trees their identity is a matter over and above physical continuity. Even if my life had been otherwise, what would make it one life could still be psychological continuity: the problem is not obvious at all.

A similar case can be presented in response to Garrett's argument. True enough, there is no physical or psychological grounding for the difference of identity between x in u and y in v, but what has that got to do with my identity over my life? Our concern is what makes y in v remain y in v over time[4], and psychological continuity appears adequate to that task even if it is not up to explaining why y in v is not x in u.

8. Another solution

These responses to Garrett and Madell point the way to an alternative solution to the dilemma that the thought-experiments appeared to set up, a solution which—like Garrett's—attempts to take the intuitions on both sides of the dilemma seriously.

The alternative resolves the dilemma by distinguishing between two questions which the exchanges examined seem to run together, and by suggesting that Locke's thought-experiment correctly occurs in the context of an answer to the one question, and all the others in the context of answering the other.

The first question is, 'What (if anything) makes someone the same person over time?' The question relates to the issue of what constitutes *my* life—in virtue of what that person *then* is me. It relates to the moral/legal issues of responsibility, desert and commitment. The second question is, 'What (if anything) constitutes personal identity across worlds?' It asks in virtue of what that person *in that world* is me; it is a modal question concerning how things might have been or how I could be different. It relates to the issues of what kind of thing/s I am *essentially*, and whatever moral consequences that may hold.

That these two questions are distinct and may need distinct answers is hardly controversial. It is common for the two to be addressed in separate sections of textbooks.

And, for instance, David Lewis's counterpart theory makes plenty of sense as a theory of (or rather, replacement theory for) transworld identity, but would be baffling as an example of a theory of transtemporal identity.

To insist on the distinction is thus not particularly interesting. What is interesting, though, is the contention (which I wish to make) that a thoroughgoing reductionism as regards personal identity does not work—in the sense that reductionist responses are not acceptable to both questions. Where Madell and Garrett go wrong is in insisting that arguments which provide evidence that transworld identity is irreducible are providing evidence for the irreducibility of trans*temporal* identity. This is in fact just another way of expressing the points made in the previous section.

Putting it this way allows an explanation of the force of the arguments presented in the name of nonreductionism. For both Madell and Garrett are correct in what they say—my life *could* have been totally different, and x in u and y in v *cannot* be identical even though nothing (apart from each being non-identical to the other) distinguishes them. These claims capture well the thesis that transworld identity is irreducible: individuals in different worlds can be *simply* identical (or distinct)—they just are the same, not *in virtue of* any other facts about their constitution, relations or origin. But the same does not go for identity across time—nothing we have heard suggests that I retain my identity as my life continues regardless of psychological or physical connections. And the very strong intuitive appeal of Locke's prince and cobbler case is very strong evidence to the contrary.

9. Concluding

By accepting that the identity of persons over time is reducible while that of their identity across worlds is not, we are able to pay due respect to the intuitive force of the kinds of arguments that Locke and Leibniz produced. Admittedly, this is at the cost of giving up other beliefs that those philosophers held, but since both clung to a number of strange beliefs this cost is not necessarily a high one. More important for us is to know which of our beliefs we must give up or accept in taking the line I have suggested. In this final section I

would like to scan quickly over what this all means.

The consequences of accepting reductionism as regards transtemporal identity need not detain us long, since Derek Parfit has discussed that in detail in *Reasons and Persons* (Parfit, 1984: ch. 15). Parfit acknowledges that certain of our important moral institutions—like that of promising, for example—are premised on a nonreductionist view of our identity over time. Accepting reductionism obliges us to revise some moral principles by (continuing the example) placing less weight on such things as long term promises, but Parfit argues eloquently that many consequences like this are in fact morally appealing ones[5].

What it means to be a nonreductionist about identity across worlds has received no such comprehensive treatment. What seems to be the most obvious consequence, though, is that if your life could have been totally different, then none of the psychological or physical features you display is essential to your being *you*. The beliefs and projects you hold most dear, your physical appearance, sex and race could all have been otherwise. The political consequences of this do not need spelling out, and may strike you uneasily as too 'enlightenment'. But then again, what can we expect from Leibniz?

Notes

1. The references to Locke's *Essay* are to Book, chapter and section. Thus 'II,xxvii, 15' refers to Book II, chapter 27, section 15.
2. For instance, as Scheffler points out, Leibniz's notion of moral identity seems to be very close to Locke's notion of personal identity.
3. This is precisely the sort of objection Peter Unger also raises against Locke's prince and cobbler case (1990, p. 158).

4. That this is also Garrett's concern is clear from the opening paragraph of his article where he specifies that it is 'personal identity *over time*' that is in focus (GARRETT, 1991, p. 361—my italics).
5. Although I am not convinced that he is telling the whole truth—see my 'Parfit and the Russians' (BECK, 1989).

References

BECK, S. (1989) Parfit and the Russians. *Analysis*, p. 49.

CURLEY, E.M. (1982) Leibniz on Locke on personal identity in: HOOKER, M (ed.) *Leibniz: Critical and Interpretive Essays.* Minneapolis, University of Minnesota Press.

GARRETT, B. (1991) Personal identity and reductionism. *Philosophy and Phenomenological Research*, p. 51.

JOLLEY, N. (1984) *Leibniz and Locke.* Oxford, Clarendon Press.

LEIBNIZ, G. (1765) *New Essays on Human Understanding.* Translated and edited by P REMNANT and J. BENNETT, (1981) Cambridge, Cambridge University Press.

LEWIS, D. (1971) Counterparts of persons and their bodies in his *Philosophical Papers* Vol. 1. New York, Oxford University Press, 1983.

LOCKE, J. (1694) *Essay Concerning Human Understanding*, edited by P.H. NIDDITCH, (1975) Oxford, Clarendon Press.

MADELL, G. (1981) *Identity and the Self.* Edinburgh, Edinburgh University Press.

PARFIT, D. (1984) *Reasons and Persons.* Oxford, Clarendon Press.

SALMON, N. (1982) *Reference and Essence.* Oxford, Blackwell.

SCHEFFLER, S. (1976) Leibniz on personal identity and moral personality. *Studia Leibnitiana* p. 8.

UNGER, P. (1990) *Identity, Consciousness and Value.* Oxford, Clarendon Press.

WIGGINS, D. (1980) *Sameness and Substance.* Oxford, Blackwell.

Simon Beck teaches in the Department of Philosophy at the University of Natal, South Africa.