

JIM STONE

HUME ON IDENTITY: A DEFENSE

(Received 12 December, 1980)

In his classic article *Hume on Personal Identity*, Terence Penelhum charges that Hume, in maintaining that we are always mistaken in ascribing identity to persons, has made "an elementary error" and fallen victim to a conceptual "muddle".¹ Hume's error, Penelhum says, is in thinking that a succession of different objects cannot be counted as one object; Hume's muddle consists in thinking it a contradiction to say that an object is both "numerically the same" and "changed". But this line of criticism ought to make us suspicious. Hume is a great philosopher and it seems unlikely that one of the central theses of the *Treatise* is merely the result of blunders.

In this paper I will argue that Penelhum has misconstrued Hume's argument and Hume's enterprise. Hume is presenting a fundamental metaphysical problem about identity through change (a problem as old as Heracleitus), not trying to analyze the ways we talk about change, as Penelhum seems to believe. In Sections I and II I will show that Hume's disturbing conclusion about personal identity does not depend upon either of the mistaken theses that Penelhum attributes to him. In Section III I explicate Hume's argument which, I argue, raises problems so serious and so general that Hume's own positive account of identity falls prey to them.

I

Penelhum observes that Hume defends the conclusion that persons do not persist through change by advancing a general thesis about identity. Penelhum writes

Hume's thesis turns on one central point, and stands or falls with it. This point is his contention that it is, "to a more accurate method of thinking", a confusion to call an object that changes the *same*. The "idea of identity or sameness" is the idea of an object that persists *without* changing. The fact that the parts of a changing thing may be related to one another does not, after all, alter the further fact that they do change; so in this case we do *not* have identity or sameness, and it must therefore be due to some ingrained tendency of the mind that we talk as we do.²

It is plain that persons (*qua* bundles of perceptions or whatever they may be) are objects that change. Hence, if Hume's contention about identity is true, it is a mistake to ascribe identity to persons. Penelhum believes that Hume's contention is plainly wrong. He quotes the passage in which it is stated

We have a distinct idea of an object, that remains invariable and uninterrupted thro' a suppose'd variation of time; and this idea we call that of *identity* or *sameness*. We have also a distinct idea of several different objects existing in succession, and connected together by a close relation; and this to an accurate view affords as perfect a notion of *diversity*, as if there was no manner of relation among the objects. But tho' these two ideas of identity, and a succession of related objects be in themselves perfectly distinct, and even contrary, yet 'tis certain, that in our common way of thinking, they are generally confounded with each other.³

Penelhum continues

It is not hard to find his error here. What he is saying is that since we could call something the same for a given length of time when it continued without any alteration, and since we would say that a succession of objects was a collection or number or series of objects, it would obviously be a contradiction to say that in the latter case we would have *one* object. ...Let us call the unchanging single object *X*. *X*, we would say, is the same throughout. Let us call our succession of distinct but related objects *A, B, C, D, E, F*, etc. Here, if we count, we obviously have several, not one. But we can quite easily produce a class-name for the series of them, say \emptyset , such that a \emptyset is, by definition, any group of thing like *A, B, C, D, E, F*, etc. So there would be no contradiction in saying there are six objects and one \emptyset ; this is what a \emptyset is.⁴

According to Penelhum, Hume has missed something obvious: a succession of different objects is *one* thing. So "there is no contradiction in saying that certain kinds of things are composed of a succession of parts and yet are each one thing"⁵. But this is a point that Hume could accept. For Hume's point is *not* that it is a contradiction to count a succession of objects as one object. Rather, Hume is maintaining that a mere succession of different objects affords us a perfect idea of diversity in that 1. it contains nothing that persists through time and 2. the existence of an ongoing diversity of objects in consistent with a universe in which nothing *substantial* persists.

Imagine a universe in which nothing persists, in which "nothing abides, all things flow". There are some trivial instances: the universe in which nothing exists, the universe consisting of one instantaneous event, and so on. But there is also the universe consisting wholly of numerically different instantaneous existents following one another in time. I think it is plain that this universe provides an idea of complete temporal diversity if we have one at all. What would we have to imagine to suppose a universe in which something

persists? There would have to be an object like a house, a tree, a stone, a particle of matter, which exists at two moments in time. These objects all belong to a class of entities traditionally called *substances* or *substantial objects* and, while I do not intend these terms to suggest any particular metaphysical theory about the nature of substantial objects, I will observe that they all have at least one necessary feature in common; namely, each can exist in its entirety at a point in time.⁶

A house, for example, can exist *as a whole* at a time *t* in that the world can contain an entire house at *t*. Of course a substantial object might be incomplete at a point in time; a house might exist without a roof at *t* if the roof is being replaced. But then it is an incomplete whole, incomplete in the sense that it could be complete at that point in time but in fact is not. By contrast the world cannot contain an entire theme of music at a point in time, because the parts that make up an entire theme exist at different times. Of course, for the same reason, the world cannot contain the entire *career* of a house at a point in time. But this only serves to distinguish the career of a house, which cannot exist as a whole at a point in time, from a house, which can. For the parts that make up a whole house (the roof, the walls, the floor) exist contemporaneously; the parts (or stages) that make up the career of the house do not.

A substantial object, I submit, exists as a whole (complete or incomplete) at every point in its career. Our idea of complete temporal diversity is of a universe in which there are no persisting *substantial* objects. It follows that an ongoing universe in which no object that is self-identical over time exists as a whole at a particular time, is a universe characterized by complete temporal diversity.

Consider, then, the objection that there must be at least *one* persisting object in the universe we have described; namely, the succession of diverse existents itself. Surely it follows from the fact that the universe is composed of *A, B, C, D, E*, and *F* happening in succession that there is a succession *A, B, C, D, E, F* that lasts from *A* to *F*, that is self-identical over the span it lasts and so on. Therefore, it is impossible to imagine an ongoing universe in which nothing persists, for a temporal diversity of objects is itself a persisting object.

The response is that though the succession *A, B, C, D, E, F* is self-identical over time, its identity is metaphysically trivial. For a temporal succession, by its very nature, cannot exist as a whole at a point in time. A substantial

object, I am maintaining, exists as a whole at every point in its career. Therefore a succession of objects is not a substantial object. It is not a persisting or enduring thing, or, if we do not wish to quarrel about the word 'persists', it is not a persisting substantial object. But a universe in which no substantial object persists is our idea of complete diversity, a Heraclitian cosmos in which we literally cannot step in the same river twice. Therefore, the claim that the only temporally self-identical object in the world is a succession of different objects *entails* that there are no persisting substantial objects. Hence, the idea of an ongoing self-identical succession of different objects is our idea of diversity.⁷

Of course Penelhum is right: for any succession of objects we can produce a class name \emptyset such that a \emptyset is, by definition, any group of things like the succession in question. So six different notes constitute *one* melody. The trouble is that if a \emptyset is, by definition, a succession of different objects then a \emptyset is not a substantial object; \emptyset s will not persist through time (in my technical sense). Of course we can count the many as one, but counting does not change the nature of the counted; it does not produce relations of identity where they did not exist before. *One* ongoing diversity is still a mere diversity. Whether we count one or many may be a function of the class-name under which we are counting. But whether we are actually presented with a persisting object, Hume maintains, is not.

II

Penelhum charges Hume with another "closely related" mistake (which he characterizes as a 'muddle'); namely, "thinking that for anything to be entitled to be called 'the same' it has to remain *unchanged* from one period to the next"⁸. Penelhum writes

The only reason for saying that something is numerically different (something else, that is) when a change occurs, is if it is by definition an unchanging thing. ...But in the case of most things, the words we use to talk about them are words the meanings of which allow us or require us to continue to use them throughout certain changes, though not, of course, *any* changes. What kind of changes can occur without our having to say that the thing has ceased to exist and given place to something else depends on what *kind* of thing we are talking about. To know what such changes are is part of what it is to know the meaning of the class-terms for that sort of object. A house, or a person, is something which admits of many changes before we would say it had ceased to exist. To know what these changes are is to know in part at least, what the words 'house' and 'person' mean.⁹

He concludes

The rejoinder to Hume, then, consists simply in saying that the pairs of expressions, (a) “numerically the same” and “containing many parts” and (b) “numerically the same” and “changed”, are not pairs of contradictories. So we have not made a mistake in saying that a succession of related objects may form a unit of a certain kind, or that the same thing may undergo radical changes.¹⁰

But this seems to miss Hume’s point. Hume admits that the ascription of identity through change does not involve “a breach of the propriety of language”.¹¹ He can accept the criticism that the linguistic conventions governing the use of class terms often allow the continued application of a class-term through change, that according to these conventions “changed” and “numerically the same” are not contradictories. Hume is not maintaining that the ascription of identity through change is ruled out on account of *linguistic conventions*. His point is that wherever there is change we cannot truly ascribe persistence because there is nothing to ascribe persistence to. This isn’t a fact about language; it is a fact about the world. Hume’s point is consistent with the fact that our linguistic conventions allow (or even require) us to apply class-terms through change. For the fact that our conventions allow or require this does not entail that any of these ascriptions are true.¹² Hume is not doing linguistic analysis as Penelhum maintains;¹³ he is doing metaphysics.

On Hume’s account it is no surprise that our conventions should have the features Penelhum mentions. For if we are subject to a powerful psychological tendency to ascribe persistence where nothing persists this is bound to affect our language. Our conventions are far more likely to reflect our practices than our philosophical insights. But we have only to press a little to see that what is allowed by our conventions is ruled out by reality. A person, for example, is a substantial object. Our concept of a person may allow us to say that a person persists through various changes. But what more is this changing person than a succession of related but different objects (e.g. a bundle of perceptions) in which nothing persists? So our ascription of identity is mistaken regardless of what our concepts allow unless, of course, we defend it by positing a metaphysical fiction.

III

We are now in a position to appreciate the force of Hume’s argument. For suppose that an object changes. Suppose for example that a part of my automobile is replaced. It follows that there is a collection of parts *P* that consti-

tutes a temporal stage of my car such that P is not identical with the collection O that existed before the part was replaced. Now my auto is a substantial object, hence, it exists as a whole (complete or incomplete) at every point of its career. Therefore if my auto (call it 'A') is self-identical over time there is an object A that exists as a whole at t_1 and an object B which exists as a whole at t_2 such that A equals B . But if my auto is merely the succession O, P it follows that my auto is not self-identical over time. For if $A = O$, then A is not identical with P (as $O \neq P$). And the succession O, P does not exist as a whole at any point in time. Therefore, if my auto persists through time it is not *merely* the succession O, P ; it must be something more.¹⁴

But what is it then? All that we are acquainted with, in fact, is a succession of collection of parts. Therefore if my auto is temporally self-identical it must be more than what we are acquainted with. To protect our ascription of identity we must "feign some new and unintelligible principle, that connects the objects together", we must "imagine something unknown and mysterious connecting the parts...".¹⁵ That is, our ascription is attended by a propensity to produce a metaphysical fiction of self or substance, something unknowable and unchanging which is the *real* object, a fiction that both Penelhum and Hume agree is unsound.¹⁶

Hume is presenting us with a profound dilemma: Either we must admit that nothing persists through change or we must posit an unchanging metaphysical substance to be the subject of persistence. For change reduces an ongoing thing to a series of *different* objects, and a succession is not itself a persisting thing. This argument does not depend on the view that a succession of different objects cannot be counted as one object, nor does it depend on the thesis that our conventions do not allow us to apply class-terms through change. Indeed the main problem with this argument may be that it is too forceful; it may undermine Hume's own theory that the idea of identity is of an object that "remains invariable and uninterrupted thro' a suppose'd variation in time". For if every change of qualities in an object marks a new object-stage, why shouldn't change of temporal properties divide an object into stages too? So the house has a temporal stage that exists at t_1 (but not at t_2) and another stage that exists at t_2 (but not at t_1). But then an uninterrupted invariable object is reduced to a succession of different but related objects by the mere passage of time, each point in time corresponding to a stage of the object. If so the idea of an object that "remains invariable and uninterrupted thro' a suppose'd variation of time" is, to an accurate view, an idea of diversity, not of sameness or identity.

Hume might defend his theory by denying that mere change of temporal properties reduces an object to a series of different objects. But this exception seems unjustified on the face of it; if the slightest change divides an object into different stages, change of temporal properties will serve as well as any. Another strategy is this: When Hume says the idea of identity is of an object “that remains invariable and uninterrupted thro’ a suppose’d variation of time” we might interpret this to mean an object that does not change any of its properties ever, including temporal properties. Such an object would have the property of existing at t_1 , say, for its entire career. So at t_2, t_3, t_4, \dots the invariable object would have the property of existing at t_1 (tenseless present). We may say that an object is *strictly invariable* when it has the same set of properties at every point in its career. Strict invariability *does* seem to be a necessary feature of persistence. For, according to the Indiscernability of Identicals, if any objects x and y are identical, then x and y share *all* their properties. The idea of temporal persistence, therefore, must be the idea of a substantial object that remains strictly invariable through a supposed variation of time; such an object will not suffer change of temporal properties.

This strategy has consequences that Hume would have difficulty accepting. First, once we allow that a strictly invariable object can have *some* dated properties (e.g. ‘exists at t_1 ’) which it keeps throughout its career, it is difficult to see why it should not have dated qualitative properties as well, for example, ‘is red at t_1 ’, which it also keeps through its entire career. So if \emptyset is strictly invariable and \emptyset is red at t_1 and white at t_2 , then at t_2 \emptyset has the property of being red at t_1 and at t_1 \emptyset has the property of being white at t_2 . In this way \emptyset will always enjoy the same set of qualitative properties. But at the same time, if \emptyset is red at t_1 and not red at t_2 then \emptyset changes, for this is simply what it is for an object to undergo qualitative change. So a strictly invariable object can undergo qualitative change. Therefore, if identity is strict invariability then identity is compatible with qualitative change.

Second, even if a substantial object is strictly invariable it will *still* be divided a series of stages by the passage of time. For at t_1 there will be a *stage* of \emptyset that exists at t_1 but not at t_2 , and vice versa, even though \emptyset itself exists at both t_1 and t_2 . But then \emptyset is not identical with any one of its temporal stages. Nor is it identical with the succession of its stages. For as \emptyset is a substantial object it exists as a whole at every point in its career, but the succession of its stages does not. It follows that a strictly invariable object is not merely the succession of its stages. It must be something more.

But what more could it be than an unchanging substance, something “unknown and mysterious, connecting the parts, beside their relation”. If we have the idea of identity at all it appears to be the idea of what Hume would call a metaphysical fiction. Every true ascription of persistence, on this account, involves an ontology that Hume rejects in the *Treatise*.¹⁷

University of Colorado

NOTES

¹ T. Penelhum, ‘Hume on personal identity’, *The Philosophical Review* LXIV (1955), pp. 571–589. Penelhum defends his paper against the criticism that Hume does *not* maintain that we are always mistaken in ascribing persistence to changing objects in T. Penelhum’s ‘Hume’s theory of the self revised’, *Dialogue* (1975), pp. 389–409. In this paper I accept Penelhum’s exposition of Hume on this point and defend Hume against Penelhum’s objections. My references are all to the original article; all italics in quotations are in the original.

² *Ibid.*, p. 576.

³ *Ibid.*, p. 579. Also, David Hume, *A Treatise of Human Nature*, Selby-Bigge edition (Oxford University Press, 1978), p. 253.

⁴ Penelhum, pp. 579–580.

⁵ *Ibid.*, p. 580.

⁶ This is a necessary condition for something qualifying as a substantial object, but it is not sufficient. An area of empty space exists in its entirety at a point in time but is not a substantial object. The specification of sufficient conditions is beyond the scope of this paper.

⁷ Henceforth, for ease of expression, I will stipulate that an object *persists* only if it is a substantial object. So some temporally identical objects will not *persist* e.g. successions. I think that this limitation is in fact implied by our concept of persistence but this point is not essential to my argument.

⁸ Penelhum, p. 580.

⁹ *Ibid.*, pp. 580–581.

¹⁰ *Ibid.*, p. 581.

¹¹ Hume, p. 258.

¹² A proposition of the form ‘ \emptyset s persist through changes like x , y , z ’ may be a conventional truth. But it doesn’t follow that when we call something a \emptyset there is anything there that persists through change.

¹³ Penelhum, p. 579.

¹⁴ The objection that the auto is identical with the subset of parts that belongs to both O and P (call it ‘ U ’) will not do. For U will cease to exist as soon as one of its elements is replaced but the auto will not. For similar reasons we cannot identify the auto with any element of U .

¹⁵ Hume, p. 254.

¹⁶ Here ‘substance’ *does* connote a metaphysical theory; namely, that substantial objects are really unchanging objects that ‘underlie’ the stages we perceive.

¹⁷ My thanks to John Nelson, Wes Morriston, Bill Prior, and especially to Arthur Millman for suggestions and criticisms of this paper.