Recent Work on Identity Over Time

Theodore Sider


I am now typing on a computer I bought two years ago. The computer I bought is identical to the computer on which I type. My computer persists over time.

Let us divide our subject matter in two. There is first the question of criteria of identity, the conditions governing when an object of a certain kind, a computer for instance, persists until some later time. There are secondly very general questions about the nature of persistence itself. Here I include the question of temporal parts, as well as certain familiar paradoxes (e.g., the statue and the lump).

Following John Perry (1975, Introduction), let us characterize a criterion of identity over time for F’s as a way of filling in ϕ in the following schema:

Stages $S_1$ and $S_2$ belong to some continuing $F$ iff $ϕ$

Defenders of temporal parts (see below) regard $S_1$ and $S_2$ as being temporal parts of the continuing $F$; others regard $S_1$ and $S_2$ as different stages in the life history of the continuing $F$. Thus each camp can make use of Perry’s formula.

It is traditional to divide such criteria into those governing persons and those governing anything else. It is further traditional to say that the criterion of identity over time for non-persons involves spatiotemporal continuity. An excellent discussion is Eli Hirsch’s The Concept of Identity¹, which utilizes the notion of continuity under a sortal. Kind-terms, or sortals, are terms that specify what kind of or sort of thing an object is. Examples include ‘tree’, ‘car’, and ‘mountain’. Where $F$ is a sortal, Hirsch’s analysis is roughly that stages belong to the same $F$ iff they are connected by a spatiotemporally and qualitatively continuous sequence of $F$-stages. Unmodified, this analysis prohibits temporally discontinuous entities, such as a watch that is taken apart and then reassembled. Hirsch discusses the necessary modifications.

Spatiotemporal continuity analyses face a problem when applied to the persistence of matter. The literature here has been dominated by discussion of examples provided by David Armstrong (1980) and Saul Kripke (unpublished

---

¹⁰ Thanks to Tamar Szabó Gendler, Trenton Merricks, Mike Rea, and Dean Zimmerman for helpful comments.

¹ See also David Wiggins’s classics: 1967; 1980.
lectures). Consider two continuous homogeneous disks, one rotating, the other stationary. Facts about the persistence of the parts of these disks differ, yet such differences do not emerge from the facts about spatio-temporal continuity, because given the homogeneity of the disks, the spatiotemporal regions occupied by the two disks are exactly similar. One must postulate some difference between the disks to account for the difference in rotation. Some defenders of *endurance* (see below) locate the difference in the life histories of enduring matter, and reject the need for a criterion of persistence for enduring matter. It is open to defenders of temporal parts to pursue an analogous strategy of postulating a *sui generis* relation uniting the stages of a continuing portion of matter, but there have been few takers. Remaining strategies include the following three. 1. Appeal to **differences in causation between the disks** (Armstrong, 1980; Shoemaker, 1979; Zimmerman, 1997). But Zimmerman has argued that this solution requires assumptions about causation and persistence that many will not be able to accept. 2. Zimmerman also points out difficulties with the next solution: 2. Postulate some unexpected qualitative feature of matter that distinguishes the disks, for example **non-Russellian velocities** (Tooley, 1988), **irreducible vector quantities** (Robinson, 1989), or **irreducible relations between temporal parts** (Hawley, 1999). The principle worry here is that the postulated “heavy-duty” ontology or ideology may be incompatible with certain reductionist agendas (although David Lewis (1986*, introduction*) argues that since the postulation is only required in worlds with homogeneous matter, the reductionist agenda could still be pursued in the actual world. 3. Appeal to **extrinsic facts about the disks** (Sider, 2001, chapter 6).

Next there is personal identity. 4. (The discussion here will be relatively brief. For a fuller discussion see James Baillie’s “Recent Work on Personal Identity” in the 1993 volume of this journal.) “Non-criterialists” reject the need for criteria of persistence for persons. Certain substance dualists are included here, though this approach is unpopular (although see Richard Swinburne’s half of Shoemaker and Swinburne (1984)). Non-dualist non-criterialism is also possible, though uncommon. 5. More typical are accounts of personal identity

---

3 See also Lewis (1999).
5 Merricks (1998) (this discussion is not particularly focused on personal identity); Lowe
in terms of either psychological or physical continuity.

The classic statement of the psychological approach is in Locke’s *Essay Concerning Human Understanding*. Influential contemporary defenses include Shoemaker’s half of Shoemaker and Swinburne (1984) and part three of Derek Parfit’s landmark book *Reasons and Persons*. The physical view has seen a recent resurgence. Its defenders include Johnston (1987a, 1992b, 1997); Olson (1997b); Thomson (1997); van Inwagen (1990b); Williams (1973). There are also hybrid positions that blend facets of the two approaches, most notably those defended by Robert Nozick and Peter Unger.

According to the psychological view, person stages belong to some continuing person iff they stand in the relation of psychological continuity. (One simple way of spelling this out: stages are psychologically continuous iff they are connected by a series of stages $S_1 \ldots S_n$ such that each $S_{i+1}$ contains apparent memories of events occurring to $S_i$.) Some argue that a causal connection is also a requirement. For suppose that a person is annihilated by a demon; and imagine further that a second demon decides, by chance, to create a person who happens to be an exact duplicate, memories and all. If the second demon’s act of creation exactly coincides spatiotemporally with the first demon’s act of destruction, we will have a sequence of person stages satisfying the requirement laid out above. And yet many intuit that the original person is nevertheless destroyed (Armstrong, 1980; Shoemaker, 1979).

Physical theories deny the relevance of psychology to personal identity, and say that person stages belong to the same person iff they are each stages of the same continuing body. Bodies must then be given some other criterion of persistence, perhaps the same criterion as obeyed by non-persons generally. A plausible variant is to use, not the entire body, but rather the brain (or whatever is the realizer of distinctive psychological capacities), as the determiner of personal identity.

Psychological theories are supported, and physical theories criticized, with the contemporary versions of Locke’s example of the prince and the cobbler. A representative example: scientists attempt to save $S$ from an incurable disease

---

6More recent work of Shoemaker’s on this topic includes his 1999.

7 A distinguished set of discussions of this volume are collected in Dancy (1997). More recent defenders include Rovane (1998); Schechtman (1996).

8 Nozick’s (1981) “closest continuer theory” allows both psychological and physical continuity to be relevant. Unger (1992) requires physical preservation of the brain plus preservation of “core psychology”.

3
by transferring her brain pattern into the brain of a previously mindless clone. S’s old body then dies, but the newly animated clone survives; call this person S’. In such examples, physical and psychological continuity come apart. Physical theories predict that S does not survive, whereas psychological theories identify S with S’. The identification of S with S’ matches our intuitions, the psychological theorists say. These intuitions may be bolstered by noting that we would blame S’ for the crimes of S, and that if thrust into this situation as S, we would fear any predicted future pains of S’. (But note that Bernard Williams (1970) has influentially argued that our intuitions about these matters are not univocal.9)

Defenders of the physical view, for example Thomson (1997); van Inwagen (1996), have directly challenged the coherence of the psychological view on general ontological grounds. A person must follow her psychology, says the psychological theorist; but clearly her body does not. Persons thus cannot be identified with human bodies. But this seems to threaten a materialist ontology, unless a metaphysic of temporal parts is accepted. Moreover, once persons and their bodies are distinguished, further puzzles follow, as van Inwagen (1990b, footnote 45) and Olson (1997b) have pointed out.10 Certain mental properties presumably supervene on momentary physical properties of the subject; but the body and the person will share all momentary physical properties. The odd consequence is that in the same location in space we have two thinkers sharing all the same thoughts. Each thinks she is a person, but one is mistaken.

In the wake of Derek Parfit’s seminal paper “Personal Identity”11, all theorists of personal identity face difficult questions about its significance. It is commonly assumed that there are a great many value-theoretic connections that hold only between a person and herself in the future. The concern one has for one’s own future (for example the dread of a future pain) seems qualitatively different than the empathy one feels for the pain of another. Only I can be blamed for my past crimes. These connections are regularly assumed in theorizing about personal identity (for example in the brain pattern transfer argument for the psychological theory given above.) And yet the connection between persistence and these values is threatened by the case of a person dividing in two.

9 An alternate response is to question the underlying methodology of thought experiments. See Quine (1972); Wilkes (1988); Gendler (2000).

10 For critical discussion see Gendler (1999); Shoemaker (1999).

11 Williams (1956–7) is an important forerunner of the contemporary discussion of fission. See also Martin et al. (1998).
Suppose that I, Ted, am divided in two, into Ed and Fred. Suppose that Ed and Fred are each strong candidates to “be me,” and that neither is a better candidate than the other. Let this occur by brain bisection, or Star Trek transporter malfunction, or whatever. The puzzle, as posed by Parfit, is then this. We cannot say that I am identical to both Fred and Ed, for by the transitivity and symmetry of identity the absurdity that Fred=Ed would follow. Nor can we identify me with exactly one of Fred or Ed, by the symmetry of their candidacy. We seem left with the conclusion that I am identical to neither; I go out of existence upon fission. And yet this too seems odd. No one would fear fission in the same way that death is commonly feared; and no one would hesitate to blame either of the resulting persons for the crimes of the original.

Parfit’s own solution is to say that fission does indeed result in the annihilation of the original person12, but that this is not bad for me, in the way that we commonly take death to be bad. Thus, Parfit denies that my continued existence has the value-theoretic significance it is commonly taken to have. What really matters to us is that our psychological lives continue, whether in us or numerically distinct future persons. This is deeply unsettling, challenging all we believe about the significance of identity. One would have thought the value-theoretic connections are constitutive of personal identity.

The literature on this puzzle has been extensive. Some responses may be roughly categorized as follows. 1. Parfit’s position: fission is not as bad as death, and so identity does not have the value we think it has. 2. Fission is as bad as death after all (Sosa, 1990). 3. By means of “tricky” metaphysics we can preserve both the view that what matters is continuation of psychological life, and the view that personal identity has its traditional significance (Lewis, 1976; Perry, 1972; Sider, 1996). 4. Parfit is right that fission would not be as bad as death; but that is because fission is a non-actual case in which ordinary concepts of concern, fear, and the like are extended beyond their normal application. In actual cases identity has the significance we ordinarily take it to have (Johnston, 1997).

Let us turn now to the second half of our subject matter. Persistence lies at the center of a set of general questions about the ontology of material objects. These questions may be broached by consideration of a paradox. Suppose we take a lump, $L$, of clay and form a statue, $S$. After our creative activity, what

---

12 More carefully, Parfit’s claim is that it is an “empty” question whether or not I survive division. But there is a best answer we can legislate: I do not survive fission at all (1984, pp. 254–266).
is the relation between $S$ and $L$? There appear to be powerful arguments for two incompatible conclusions. Conclusion 1: $S = L$. Argument: $S$ and $L$ now occupy the same place. Moreover, they share exactly the same parts, have the same mass, velocity, and so on. How could two distinct objects fit into the same location in space, share exactly the same parts, and so on? Conclusion 2: $S \neq L$. Argument: $S$ and $L$ have different properties. Our creative activity created $S$. Before we formed $L$ into statue shape, $S$ did not exist. But $L$ did exist then. Therefore, $L$, but not $S$, has the property existing before being formed into statue shape. By Leibniz’s Law, $S$ and $L$ are distinct.

There are in fact a number of loosely related paradoxes that similarly threaten our ordinary beliefs about material objects and persistence (e.g., Tipples and Tib, the paradox of fission discussed above, the ship of Theseus.) Different ontologies of persistence may be distinguished by how they resolve these puzzles. Rea (1997) collects many of the important papers on this topic, and the introduction contains a survey of the positions one can take on coinciding objects.\textsuperscript{13}

One solution is to say that continuants 	extit{perdur}, i.e., are composed of temporal parts.\textsuperscript{14} The statue and the lump are numerically distinct, but fit into the same location in space because they share a common temporal part at the time. Indeed, the statue is a proper temporal part of the lump. My forthcoming book \textit{Four-Dimensionalism} contains a defense of this ontology, including a critical discussion of existing arguments in the literature, and new arguments for temporal parts.

One prominent argument for temporal parts is Lewis’s argument from “temporary intrinsics”, in that contemporary classic \textit{On the Plurality of Worlds}. The traditional problem of change is that changing things seem to contravene Leibniz’s Law, by instantiating incompatible properties. The glib solution is that the incompatible properties are had at distinct times. But Lewis argues that if the change occurs with respect to \textit{intrinsic} properties, the glib solution is blocked, for intrinsic properties are non-relational and hence do not hold with respect to anything, not even times.\textsuperscript{15}

\textsuperscript{13}See also Rea (1995); Sidelle (1998); Sider (2001, chapter 5).


\textsuperscript{15}Lewis (1986a, 202–204). On this argument see also: Forbes (1987); Haslanger (1989);
Sadly, many philosophers reject temporal parts (Oderberg, 1993; Rea, 1998; Thomson, 1983; van Inwagen, 1990a). What solutions to the puzzle of the statue and lump are then available? The most prominent is that of David Wiggins. If any solution deserves the label of orthodoxy it is this one. Like the temporal parts theorist, Wiggins agrees that the lump and statue are numerically distinct. But Wiggins, like the rest of the theorists to be discussed from now on, denies that objects perdure. Rather, they **endure**: they have no temporal parts, but are rather “wholly present” at every moment at which they exist. The statue and lump can nevertheless share spatial location because i) they are of different sorts, and ii) the statue is *constituted* by the lump. Wiggins then faces the question of what constitution amounts to, and how it allows sharing of spatial location. The recent literature also contains another important challenge. According to Wiggins, the statue and lump differ by having different historical (and also modal) properties: only $L$ has **existing before being formed into statue shape**. But on what is this difference based? Since $S$ and $L$ share all their momentary intrinsic properties, share the same parts, and so on, any difference in historical (or modal) properties seems to violate a plausible supervenience principle for such properties.

Another approach is to claim that $S$ and $L$ are not numerically distinct after all. This can be made out in a number of different ways. One might say that forming $L$ into statue shape does not *create* anything, but rather causes $L$ to take on the property of **statuehood**. But when *are* things created? Suppose we obtained $L$ by chemically transforming a pre-existing aggregate of matter. Was this a creation of $L$, or did the pre-existing aggregate merely acquire the property **being a lump of clay**? The least arbitrary answer to questions like this is given by the **mereological essentialist**, for example Roderick Chisholm. On

---


this view, what exists are aggregates of matter that have their parts essentially. In our statue and lump example we have a single material object, an aggregate of matter, that has existed since its parts came into existence (and, perhaps, were “stuck together”), and will continue to exist so long as those parts remain (and, perhaps, remain stuck together). On its face this view has counterintuitive consequences. If I replace a tire of my car, the resultant car is not identical with the original. Mereological essentialists respond that ordinary assertions of cross-time identity do not attribute “strict and philosophical identity”.

An alternate way of identifying $S$ and $L$ has been proposed recently by Michael Burke. On his view, the formation of the lump of clay into a statue does indeed create $S$, but this does not result in two distinct things in the same place, for the original lump of clay is destroyed. In its place there comes to exist a new lump of clay, $L$, which is identical to the resultant statue, $S$.\(^{19}\)

An even more radical solution to the puzzle is one of the upshots of probably the most influential metaphysics book of the 1990s: van Inwagen’s *Material Beings*.\(^{20}\) For independent reasons, van Inwagen argues that most of the objects of our everyday conceptual scheme do not exist. There are no such things as tables, chairs, planets, statues, or lumps of clay. (Van Inwagen makes an exception for living things, but this need not detain us.) This is not to say that ordinary assertions about these non-entities are all false. For van Inwagen, an ordinary utterance of ‘there is a book on a table’ has roughly the following truth conditions: there are some $X$s and some $Y$s, such that i) the $X$s are arranged bookwise, ii) the $Y$s are arranged tablewise, and iii) the $X$s are on the $Y$s. The values of the plural variables ‘$X$s’ and ‘$Y$s’ will be simples — objects without proper parts — whose genuine existence van Inwagen does accept. Though this fact does not feature prominently in the book, the ontology of *Material Beings* dissolves the puzzle of the statue and lump: the culprit entities do not really exist.\(^{21}\)

Yet more radical solutions have been relatively unpopular: denying one of the quasi-logical assumptions implicit in the argument, for example the transitivity of identity or the principle that identity is a two-place relation between continuants without an argument place for times or sortal predicates.\(^{22}\)


\(^{20}\) See also van Inwagen (1981). For criticism see Sider (1993); Rosenberg (1993); Horgan (1993).

\(^{21}\)This feature of the ontology gets more central billing in Merricks (2001).

\(^{22}\)For denials of the latter assumption see Chandler (1971); Geach (1997); Myro (1986);
It should be evident that the contemporary discussion of identity over time is rich and flourishing.

References


