forthcoming in *Thought: A Journal of Philosophy*

Homunculi are people too!
Lewis's definition of personhood debugged

Cody Gilmore
UC Davis

1. Introduction

In 'Survival and Identity', David Lewis (1976a) defends a psychological theory of personal identity and a stage-sharing treatment of fission and fusion.[1] Standard cases of personal fission, Lewis holds, involve exactly two people; these people share all their pre-fission stages but not their post-fission stages. They are like overlapping roads that share their initial segment (their 'trunk') but diverge at a fork. Fusion is handled in a similar way.

One nice feature of Lewis's theory is that its verdicts about fission and fusion are not *ad hoc* stipulations. They are logical consequences of the elegant 'non-circular definition of personhood' at the heart of the theory:

> (L) something is a continuant person if and only if it is a maximal R-interrelated aggregate of person-stages. That is: if and only if it is an aggregate of person-stages, each of which is R-related to all the rest (and to itself), and it is a proper part of no other such aggregate. (1976a: 22)

Another consequence of (L) is that personhood is *maximal*, where a property is maximal if and only if it is impossible that a thing and one of its proper parts both have the property.[2] But this is a bug, not a feature. As Michael Burke (2003: 112) has pointed out (though not in connection with (L)), there are counterexamples to the maximality of personhood.[3] A person could be a part of another, much larger person with a completely separate mental life. The counterexamples are obvious enough that there's already a word for them: homunculi.[4]

Admittedly, the homunculus example is exotic and probably non-actual, but this does not diminish its relevance. Lewis intends his theory to hold 'not only for the cases that arise in real life, but for all possible problem cases as well' (1976a: 22),[5] and he places great weight on cases involving adult human fission, which are hardly less exotic

---

[1] Everything that I say about fission and fusion in this paper can also be said, *mutatis mutandis*, about what Lewis calls 'longevity'.

[2] To say that x is a proper part of y is to say that x is a part of, but not identical to, y. Strictly speaking, the maximality of personhood is a consequence not of (L) but of its necessitation, which Lewis endorses, given that he offers (L) as a definition.

[3] Lewis (1983: 41) offers a definition, call it *(M)*, of 'modal continuant' that is parallel to (L), and which suffers from a parallel bug, *modulo* the fact that 'modal continuant', unlike 'person', is a technical term. Those who, unlike Lewis, wish to treat people as modal continuants will want to debug (M) in the manner I debug (L).

[4] In literature, and in 17th century preformationist hereditary theory, a homunculus is a tiny person or humanoid, often embedded within an ordinary person.

[5] In this passage Lewis is referring to his claim that the I-relation and the R-relation are coextensive, but it is clear that he takes (L) to be metaphysically necessary as well, given the role it plays in determining the extension of the I-relation, and given that he calls it a definition. On the R-relation, see section 2. On the I-relation, see note 15.

than homunculus cases. So if the homunculus case is even metaphysically possible, it is a problem for Lewis's definition.

In section 4, I note that Lewis is especially vulnerable to the counterexample, given his views on personhood, mental properties, and the extent of metaphysical possibility. I then show, in section 5, that Lewis's definition is remarkably easy to debug. A conservative repair, which retains all the virtues of (L) and introduces no new vices, is just a minor amendment away.[6]

2. Preliminaries

Three expressions in (L) require comment. (i) 'Person-stage' is primitive, but it can be glossed as 'an instantaneous or very short-lived entity that is sufficiently person-like, especially with regard to its psychological profile'. (ii) 'x is R-related to y (*simpliciter*)' can defined as 'x and y are each person-stages, and either (a) x is to some extent under the intentional control of y and y is accessible in memory to x, or (b) *vice versa*, or (c) x=y'.[7] R-relatedness, according to Lewis, is a relation of psychological connectedness that is reflexive (over the domain of person-stages) and symmetric but not necessarily transitive. (iii) 'Aggregate' is defined as 'mereological sum':[8] x is a mereological sum of plurality yy if and only if each of yy is a part of x and each part of x overlaps (shares a part with) at least one of yy.[9] I will use 'aggregate' and 'sum' interchangeably. I assume that parthood is reflexive and transitive and that each plurality has exactly one sum.[10]

3. The bug

The maximality clause in (L) is motivated by the following thought. The proper temporal part[11] of (e.g.) Lincoln that runs from the beginning of 1850 to the end of 1859, call it *1850s-Lincoln*, is an R-interrelated aggregate of person-stages, but it is not a person. So 'person' must not be defined simply as 'R-interrelated aggregate of person-stages'. In an ordinary case, a proper temporal part of a person is an R-interrelated aggregate of person-stages but is not a person.

One natural alternative is (L). Since 1850s-Lincoln is a proper part of another R-interrelated aggregate of person-stages (e.g., Lincoln himself), (L) counts 1850s-Lincoln as a non-person. So far, so good.

However, a harder case for (L) is this:

---

[6] Hudson (2001: 144) proposes a sophisticated account of what it is to be a *human* person (as opposed to a person *simpliciter*). A variant of the homunculus case serves as a counterexample to Hudson's account if it is possible that both of the people in the case are human people (of which I am uncertain).

[7] This is my paraphrase of Lewis (1976a: 23-24).

[8] Lewis (1976a: 39, note 4) says that 'mereological sum' is his preferred interpretation of 'aggregate' but that other interpretations would work as well. In other work he shows no hesitation about treating people as mereological sums of stages.

[9] I use 'xx', 'yy', etc., as plural variables, which range over pluralities, and I treat 'is one of' as primitive. To say that xx are *among* yy is to say that, for each z, if z is one xx, then z is one of yy. To say that xx *are other than* yy is to say that there is some z such that either z is one of xx but not one of yy, or z is one of yy but not one of xx. On plural logic, see Lewis (1991) and Linnebo (2016).

[10] See Lewis (1991), Hovda (2009), and Varzi (2016).

[11] The proper temporal parts of a thing are proper parts of the thing. On the definition of 'temporal part', see Sider (2001: 55-62).

Allyson. Allyson is a sum of *aa*, which are some instantaneous, R-interrelated person-stages, and which are not among any other R-interrelated plurality of person-stages. Indeed, none of *aa* is R-related to anything that is not one of *aa*.

Allyson is physically and psychologically just like an ordinary human being who lives for 95 years. However, Allyson is a proper part of a vastly larger conscious being.

This larger being, call it *Zebulon*, is a sum of some (extremely large) instantaneous person-stages, *zz*, which are R-interrelated. Zebulon is not a proper part of any other R-interrelated aggregate of person-stages, and none of *zz* is R-related to anything that is not one of *zz*. None of *aa* is one of *zz*, though each part of each of *aa* overlaps at least one of *zz*. (In other words, no part of any of *aa* is disjoint from each of *zz*.[12]) Many of *zz* have parts that are disjoint from each of *aa*.

Zebulon and Allyson both have whatever psychological features (self-consciousness, higher-order desires, etc.) are deemed relevant to personhood. But Zebulon and Allyson are as separate psychologically as any two people ever have been. Neither of them is aware of or remembers any experience had by the other, and neither is under the intentional control of the other. Allyson is usually happy and thinks mainly about numbers and sports. Zebulon is usually depressed and thinks about mainly about literature.

The details can be specified in different ways. Perhaps Zebulon is the 'system' composed of a Searle (1980)-style Chinese Room and its contents, including Allyson; and this system is implementing a mentality-generating computer program. Perhaps Zebulon is a Block (1980)-style Nation of China, and Allyson is one of its human neuron-surrogates. (Searle and Block would in both cases deny that Zebulon is conscious, but many disagree.) Perhaps Zebulon is an organism-like entity, trillions of light years across, whose cell-analogues are composed of super-clusters of galaxies, one of which is inhabited by Allyson.

Details aside, the key is this. Zebulon is an R-interrelated aggregate of person-stages and is not a proper part of any other such aggregate. So, according to (L), he is a person. Allyson is a proper part of an R-interrelated aggregate of person-stages, namely Zebulon. So, according to (L), she is not a person. In fact, however, both Allyson and Zebulon are people, at least given a broadly psychological approach to personhood, which I will not challenge here. So (L) is incorrect. Allyson is a counterexample to (L).

4. Escape routes blocked

In this section I mention a few strategies for resisting the example, and I note, in a mostly *ad hominem* fashion, that they are not open to Lewisians, given Lewis's views on related topics.

Route 1. One might deny that the appropriate network of causal relations could be implemented on a super-human scale. For example, one might deny that a billion people could use radios to signal to one another in a way that mirrors the neuronal firing patterns in the brain of a conscious human being, as in Block's Nation of China case.

[12] 'x is disjoint from y' means 'x does not overlap y'.

However, even if this is nomically impossible, presumably it is still metaphysically possible, which is the relevant type of modality. Given Lewis's liberal views about what is metaphysically possible (including backward causation and backward time travel (1976b), causation via magic spells (1983:76), person-stages appearing *ex nihilo* and vanishing into thin air (1983: 76), interpenetrating, non-interacting systems of matter and energy (1986: 72), and alien properties (1986: 92)), Lewisians ought to grant that there is a metaphysically possible world at which the relevant causal facts obtain.

Route 2. One might grant the possibility of a world in which these causal facts obtain but deny that the 'Zebulon object' has mental properties in such a world, and so deny that it counts as a person. Perhaps, e.g., mental properties can only arise from causal interactions between *neurons*, whereas Zebulon's mental properties, if such there be, would arise only from causal interactions between non-neurons. Lewis, however, reduces mental properties to the occupants of certain causal roles (e.g., 1986a: 106) and endorses the possibility of entities that have mental properties despite lacking neurons (1983: 123). According to his theory of mind, the relevant causal facts[13] suffice for Zebulon's having mental properties.

Route 3. One might deny that Allyson's stages, *aa*, are person-stages, on the grounds that (i) *being a person-stage* is maximal, and (ii) each of *aa* is a proper part of some person-stage, one of *zz*.

This suggestion fails for two reasons. First, the Allyson case does not entail (ii), but only that no part of any of *aa* is disjoint from each of *zz*. That is consistent with the proposition that none of *aa* is a part of any of *zz*. Perhaps, e.g., each of *aa* is instantaneous only in a certain inertial frame, F, whereas each of *zz* is instantaneous only in a different inertial frame, F*, so that each of *aa* overlaps many of *zz* but is a part of none of *zz*. Second, if (ii) were built in to the Allyson case, that case would serve as forceful counterexample to (i), the maximality of person-stage-hood.

5. Lewis's definition debugged

Here is the amended definition:

> (L*) x is a continuant person if and only if x is an aggregate of some person-stages, xx, each of which is R-related to all the rest and to itself, and xx are *maximal with respect to R-interrelatedness*, in the sense they are not among some other plurality of person-stages, yy, each of which is R-related to all the rest and to itself.[14]

Allyson is an aggregate of some R-interrelated person-stages, *aa*, that are not among any other R-interrelated person-stages. So, although Allyson is a *proper part* of another R-

[13] Lewis's theory of mental properties (1983: 122-132) requires that we add certain assumptions about which properties occupy which causal roles *in the population to which Zebulon belongs*. Consider it done.

[14] The use of plural logic is dispensable. (L*) can be stated terms of sets: x is a person if and only if x is an aggregate of some set S of person-stages, where each member of S is R-related to itself and to every other member of S, and S is maximal with respect to R-interrelatedness, in the sense that S is not a subset of some other set S* of person-stages such that each member of S* is R-related to itself and to every other member of S*. (To say that x is an aggregate/sum of a set S is to say that each member of S is a part of x and each part of x overlaps at least one member of S.)

interrelated aggregate of person-stages (Zebulon), (L*) counts her as a person. The key fact is that *aa* are not among *zz*, nor are *aa* among any other R-interrelated plurality of person-stages. Indeed, none of *aa* is identical to *any* of *zz*. (Each of *aa*, but none of *zz*, is human-sized.) [15]

(L*) yields the same result as (L) when applied to Zebulon. Zebulon is an R-interrelated aggregate of person-stages, *zz*, which are not among any other R-interrelated plurality of person-stages. So (L*), like (L), counts Zebulon as a person, as desired.

I leave it to the reader to check that (L*) yields the same verdicts on ordinary cases, and on standard cases of fission and fusion, as does (L). So, since (L*) retains the virtues of (L) and avoids one of its vices without adding any new ones, I recommend that Lewisians replace (L) with (L*).[16]

References

Block, N. 1980. 'Troubles with Functionalism', in N. Block, ed., *Readings in the Philosophy of Psychology, vol. 1* (Cambridge, MA: Harvard University Press), pp. 268-305.

Burke, M. 2003. 'Is my head a person?', in Klaus Petrus, ed., *On Human Persons* (Frankfurt: Ontos Verlag), pp. 107-125.

Hovda, P. 2009. 'What Is Classical Mereology?', *Journal of Philosophical Logic*, 38: 55–82.

[15] Once we have replaced (L) with (L*), Lewis's account of the I-relation needs adjustment. Like the R-relation, the I-relation holds between person-stages, not continuant persons. But the I-relation is defined in a way that is neutral with respect to the debate between psychological and, e.g., bodily theories of personal identity over time: 'S_1 and S_2 are I-related . . . if and only if there is some one continuant person of whom both S_1 and S_2 are stages' (1976a: 23). Call that definition *(DI)*. Lewis regards it as a common sense platitude that *being I-related to some future person-stage is what matters in survival*; this is his surrogate for the standard view that identity is what matters in survival, and the main job the I-relation is to allow him to formulate this surrogate. He regards it as a controversial philosophical hypothesis – superficially in tension with the above platitude – that *being R-related to some future person-stage is what matters in survival*. And he holds that, given his psychological theory of persons, encapsulated by (L), the platitude and the controversial hypothesis are both true.

However, if 'x is a person-stage of y' is defined as 'x is a person-stage, y is a continuant person, and x is a part of y' – call that definition *(DS)* – then (DI), above, yields results that conflict with other things Lewis says about the I-relation. For example, given the facts of the case, (DS) and (DI) yield the result that (*) each of Alysson's person-stages is I-related to each of Zebulon's person-stages, despite the fact that they are not R-related. But Lewis writes, 'I claim that *any stage is I-related and R-related to exactly the same stages*.' (1976a: 22, italics original). To avoid (*), one has three main options. First, one could replace (DI) with (DI*): S_1 and S_2 are I-related if and only if each of them is one of some plurality xx of person-stages, no two of which overlap, and which have a person as a sum. (DI*) handles the case of Allyson and Zebulon, but, for what it's worth, it has trouble with cases in which a person, P, travels backward in time, shrinks, and moves around in such a way that some small, 'older' person-stages of P are proper parts of the larger, 'younger' person-stages of P. See Kleinschmidt (2011) for similar cases not involving people. Second, one could define 'is I-related to' as 'is R-related to', vindicating by fiat Lewis's claim that 'the I-relation is the R-relation' (1976a: 22), though this would undermine Lewis's use of the I-relation to capture, in neutral terms, the common sense platitude about what matters in survival. Third, one could reject (DS). One might take the dyadic predicate 'is a stage of' as primitive, in which case one could then define the monadic predicate 'is a stage' in terms of it, as 'is a stage of something'.

[16] I am grateful to the Immortality Project, directed by John Martin Fischer and funded by the John Templeton Foundation, for supporting the research of which this paper is an outgrowth. Thanks also to I-Sen Chen, T. Scott Dixon, and Hud Hudson for helpful feedback. Special thanks to an anonymous referee for insightful comments that made the paper better, and shorter, than it was originally.

Hudson, H. 2001. *A Materialist Metaphysics of the Human Person* (Ithaca: Cornell University Press).
Kleinschmidt, S. 2011. 'Multilocation and Mereology', in J. Hawthorne, ed., *Philosophical Perspectives*, 25: 253–276.
Lewis, D. 1976a. 'Survival and Identity', in A. O. Rorty, ed., *The Identities of Persons* (Berkeley, CA: The University of California Press), pp. 17-40.
Lewis, D. 1976b. 'The Paradoxes of Time Travel', *American Philosophical Quarterly* 13: 145-152.
Lewis, D. 1983. *Philosophical Papers, Volume I* (Oxford: Oxford University Press).
Lewis, D. 1986a. *On the Plurality of Worlds* (Oxford: Blackwell).
Lewis, D. 1986b. *Philosophical Papers, Volume II* (Oxford: Oxford University Press).
Lewis, D. 1991. *Parts of Classes* (Oxford: Blackwell).
Linnebo, Ø. 2014. 'Plural Quantification', *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2014/entries/plural-quant/>.
Searle, J. 1980. 'Minds, Brains, and Programs', *Behavioral and Brain Sciences* 3: 417-424.
Sider, T. 2001. *Four Dimensionalism: An Ontology of Persistence and Time* (Oxford: Oxford University Press).
Varzi, A. 2016. 'Mereology', *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2016/entries/mereology/>.