

Stage Theory and the Personite Problem

Forthcoming in Analysis – please cite published version

Abstract: Mark Johnston has recently argued that four-dimensionalist theories of persistence are incompatible with some of our most basic ethical and prudential principles. I argue that although Johnston's arguments succeed on a worm-theoretic account of persistence, they fail on a stage-theoretic account. So much the worse, I conclude, for the worm theory.

1. Introduction

Mark Johnston (2016, 2017) has recently argued that four-dimensionalist theories of persistence have disastrous consequences for practical ethics. The problem, he argues, is that such theories are committed to the existence of *personites*, shorter-lived objects coincident with people; that personites have moral status on any plausible theory of how things acquire such status; and that this consequence is incompatible with even our most basic ethical and prudential assumptions. Making short-term personal sacrifices for long-term personal gains turns out to be deeply immoral, for example, because the sacrifices will also have to be endured by countless personites who will go out of existence before the gains are realized. Keeping one's promises is equally immoral, because it forces countless personites who came into existence after the promise is made to do something they have no duty to do. And the intimacies of close friendship involve impertinence on a massive scale, because they are imposed on a host of recently arrived personites that had nothing to do with the background of established consent that makes such intimacies ethically acceptable.

I agree with Johnston that four dimensionalism is committed to the existence of personites. I also agree with him that *if* personites have moral status, much of our ethical and practical reasoning is wrong. My focus will be on Johnston's argument that personites have moral

status. I will show that his argument only works on a *perdurantist* or *worm-theoretic* account of persistence, according to which people instantiate tensed properties in virtue of having temporal parts at past and future times, but fails on an *exdurantist* or *stage-theoretic* account of persistence, according to which people instantiate tensed properties in virtue of having temporal *counterparts* at past and future times. Far from being a *reductio* of four dimensionalism, then, the personite problem should really be interpreted as a novel argument for one of its variants over the other.

2. Preliminaries

Following Sider (2001), let's start by defining some standard terminology. Our basic undefined concept shall be that of *parthood-at-t*, which is reflexive and transitive. Two objects *overlap* at a time iff there is something which is a part of both of them at that time. *x* is a *fusion* at *t* of class *S* iff (i) every member of *S* is a part of *x* at *t*, and (ii) every part of *x* at *t* overlaps at *t* some member of *S*. *x* and *y* *coincide* at *t* iff they overlap at *t* exactly the same objects. Finally, *x* is an *instantaneous temporal part* of *y* at *t* iff (i) *x* exists at, but only at, *t*, (ii) *x* is part of *y* at *t*, and (iii) *x* overlaps at *t* everything that is part of *y* at *t*.

Four dimensionalism is the view that every object has an instantaneous temporal part at every time at which it exists. The *worm theory* and the *stage theory* are both four-dimensionalist theories. They both agree, roughly speaking, on what there is, and on which things are parts of which other things. Where they disagree is on the semantics of tense and the referents of ordinary terms. Let's say that something is a *person-stage* iff it is an instantaneous temporal part of a person, and let *R* be that relation between person-stages that matters to personal identity through time (plausible candidates include psychological continuity, bodily continuity, and so on). According to the worm theorist, people are *maximal R-interrelated*

fusions of person-stages.¹ What *was* or *will* be true of me at t is grounded in what is true *simpliciter* of an instantaneous temporal part of me that exists, atemporally speaking, at t . It's true that I will be in France later this year, for example, in virtue of the fact that there is a temporal part of me in the future that is (atemporally speaking) in France. According to the stage theorist, by contrast, *people are person-stages*. What *was* or *will* be true of me at t is grounded in what is true *simpliciter* of someone R-related to me that exists, atemporally speaking, at t . It's true that I will be in France later this year, for example, in virtue of the fact that I am R-related to someone in the future who is (atemporally speaking) in France.²

We're now in a position to say more precisely what a personite is. Let's say that an object x is *temporally continuous* iff for all times $t_1 < t_2$, if x exists at both t_1 and t_2 , then x also exists at all times t such that $t_1 < t < t_2$. Then:

If x is a person, then y is a personite of x iff y is a temporally continuous *non-maximal* R-interrelated fusion of two or more person-stages, each of which is R-related to every instantaneous temporal part of x .³

¹ This shouldn't be interpreted as a *definition* of a person, of course – after all, person-stages were themselves defined in terms of people.

² Johnston has put it to me in conversation that the stage theory is committed to a problematically disjunctive account of existence-at-a-time. I'm not sure I agree. It's true that according to the stage theory, I *will* exist at times at which, atemporally speaking, I don't exist. But this isn't a disjunctive analysis, because the (tensed) claim that I will exist at t is not the same as the (untensed) claim that I exist at t . One might claim to find it intuitively incredible that I don't exist at times at which I will exist. But firstly, this is a quite different complaint; and secondly, I am more than a little suspicious of those who claim to have intuitions about what is atemporally the case, over and above their intuitions about what was and will be the case.

³ Note that although the stage theorist and the worm theorist disagree on what *I* am, they won't disagree, given this definition, on which things are my personites; *at least* in those cases where R is an

If four dimensionalism is correct, I have a great many personites.⁴ There is the personite which came into existence when I was born and went out of existence five years ago, for example, as well as the personite which will come into existence in four seconds time and go out of existence 1.7 seconds later. Yet although *I* am a person, both the worm theory and the stage theory imply, albeit for different reasons, that none of my personites are people. According to the worm theory, my personites aren't people because although they are R-interrelated fusions of person-stages, they aren't *maximal* such fusions – there are person-stages that are R-related to every instantaneous temporal part of my personites that aren't themselves instantaneous temporal parts of them. According to the stage theorist, by contrast, my personites aren't people simply because they aren't person-stages; rather, they are fusions of person-stages.

Notwithstanding the fact that my personites aren't people, however, Johnston thinks that they nevertheless ought to be granted moral status. The next section examines Johnston's argument for this claim. I will show that although his argument works on the worm theory, it fails on the stage theory.

3. The Personite Problem

equivalence relation. In fission and fusion cases, where R is non-transitive, matters are less straightforward – if I will undergo a fission event in the future, the stage theorist will say that I have a personite with parts in both 'branches'; the worm theorist will deny this (instead, they will insist that my personites are only made up of person-stages from one branch, although I am currently coincident with a different person whose personites are made up of person-stages from the other branch). I'll ignore this complication from now on, as it won't be relevant to the main argument of the paper.

⁴ I'm assuming here that, for any class of person-stages, there is something which is the fusion of that class. This doesn't strictly speaking follow from four dimensionalism as defined, but most four dimensionalists at least implicitly endorse something like this thesis.

Johnston's (2016: 203-4) argument for the proposition that personites have moral status proceeds as follows. Let's say that, for all possible objects x and y and possible worlds w and v , x in w is a *duplicate* of y in v iff x instantiates exactly the same intrinsic properties in w as y does in v . Then:

1. For all possible worlds w and possible objects x , if x is a person in w then x has moral status in w .
2. For all possible worlds w and v and possible objects x and y , if x in w is a duplicate of y in v , then x has moral status in w iff y has moral status in v .
3. For all personites x , there is a possible object y and possible world w such that y is a person in w and y in w is a duplicate of x in the actual world.
4. Therefore, all personites have moral status.

Let's grant premises 1 and 2 for the sake of argument, and focus instead on premise 3. Johnston motivates this premise by arguing that it is perfectly possible – i.e. not incompatible with the concept of personhood – for a person to spring into existence fully-formed (e.g. 'from a chemical swamp or...as with Athena, emerging whole from the forehead of Zeus' (2016: 204)); and moreover that it is perfectly possible for people to die suddenly at any time before the time of their actual death. From this and the worm theory it follows that, for all times $t_1 < t_2$ during my lifetime, there is a possible world w containing a person S who comes into existence at t_1 and goes out of existence at t_2 who is in all intrinsic respects exactly like I actually am between those times. By the definition of personite, then, it follows that for every one of my personites, there is a possible world w containing a person S who is a duplicate in w of my personite in the actual world. The only difference between them is that, in the actual world, there are person-stages before t_1 and after t_2 that are R-related to every instantaneous temporal part of my personite, whereas in w there are no person-stages before t_1 or after t_2 that are R-related to every instantaneous temporal part of S . These *extrinsic* differences suffice to ensure that S is a *maximal* R-interrelated fusion of person-stages and my personite is not,

and hence, on the worm theory, that S is a person and my personite is not; but such extrinsic differences, by premise 2 of Johnston's argument, cannot suffice to ensure that S, but not my personite, is worthy of moral status. Hence if S has moral status in w , my personite has moral status in the actual world.

I agree with Johnston that worm theorists are required to accept premise 3. The question is whether the same argument succeeds on a stage-theoretic account of persistence. The stage theorist will agree with the worm theorist that, for every one of my personites, there is a possible world w containing a maximal R-interrelated fusion of person-stages which is a duplicate in w of my personite in the actual world. They will agree, that is, that for every *non-maximal* R-interrelated fusion of person-stages, there is a possible duplicate of that object that is a maximal R-interrelated fusion of person-stages. Crucially, however, *they won't agree of either of these things that it's a person*, because for the stage theorist people are person-stages, and neither of these things is a person-stage. Hence it doesn't follow on the stage theory that my personites are duplicates of some possible person – they are duplicates of some possible maximal R-interrelated fusion of person-stages, but people aren't fusions of person-stages (maximal or otherwise). In short, Johnston's argument for premise 3 fails on a stage-theoretic account of persistence.

One might insist that, even if my personites aren't duplicates of a possible *person*, they are nevertheless duplicates of something with moral status, because maximal R-interrelated fusions of person-stages have moral status regardless of whether we decide to call them 'people'. But I don't see why a stage theorist should accept this premise. From the perspective of the stage-theorist, the four-dimensional objects worm theorists call 'people' are no different from, say, the fusion of me and Mick Jagger (or, more to the point, the fusion of me and my *modal* counterparts, if such things there be) – perfectly real, but not legitimate objects of moral or prudential concern, despite being composed of things which are. The fact that one is a maximal R-interrelated fusion and the other is not is neither here nor there. Indeed, one of the alleged advantages of the stage theory is precisely that extrinsic facts like whether

something is R-related to something else do not make a difference to whether either of those things is a person.

I've argued in this section that the personite problem is not a problem for *all* versions of four dimensionalism. But it's worth noting that the personite problem is arguably not a problem for *only* versions of four dimensionalism either. Consider the view Magidor (2016) calls a 'liberal endurantism', which agrees with the worm theory that people are four-dimensional objects that are *co-located* with a multitude of shorter-lived objects at every time of their existence, but just denies that these shorter-lived objects are *parts* of the person. In particular, the liberal endurantist denies that a person has an instantaneous temporal part at every time at which they exist, and hence rejects four dimensionalism, as standardly defined. Nevertheless, a similar argument to Johnston's would suffice to show that these shorter-lived objects have moral status, in virtue of being duplicates in this world of a person in some other world. In summary, the personite problem is a problem for any view according to which a person is co-located, at every time at which it exists, with a multitude of objects that are intrinsically identical to the person at that time, whether these objects turn out to be parts of the person or not. The stage theory, however, is not such a theory.

4. Objections and Replies

I will now consider some objections to my claim that the stage theory avoids the personite problem.⁵

Objection: The stage-theoretic 'solution' succeeds only in moving the bump in the carpet. After all, the stage theorist accepts that I instantiate temporal properties, notwithstanding my lack of temporal extension; so on pain of metaphysical arbitrariness it seems they must also admit the existence of a multitude of *other* instantaneous objects coincident with me, but with

⁵ This section is based on an exchange I had with Johnston at the 2018 Evans Memorial Lecture in Oxford, for which I am very grateful.

different temporal properties. Coincident with me right now, for example, is something which *has* only existed for several seconds, as well as something which *will* only exist for several seconds more. Of course, atemporally speaking none of these objects exist for more than an instant, according to the stage theorist. But they're nevertheless different objects in virtue of having different temporal counterparts. Moreover they are duplicates of me, and hence worthy of moral consideration by straightforward application of premise 2 of the argument above. All the same problems for ethics and practical reasoning then re-emerge.

Response: Rather than admitting a multiplicity of entities with different temporal counterparts, the stage theorist should instead admit a multiplicity of counterpart relations and relativize the predication of temporal properties to a choice of counterpart relation, usually determined by the conversational context (see Lewis 1986: ch.4). There is just one object coincident with me – *me* – but that object instantiates different temporal properties relative to different counterpart relations. This way, the stage theorist can dodge the charge of metaphysical arbitrariness without postulating the existence of a multiplicity of entities coincident with me.

Objection: But now things look even worse for the stage theorist. What it is practically rational for me to do now depends on what I will do in the future. But if facts about what I will do are relativized to a counterpart relation, it seems that facts about what it is practically rational for me to do now must also be relativized to a choice of counterpart relation. And that's absurd! If it's true of me that I ought to save for retirement, it's true *simpliciter*, not simply true *qua* person. *Mutatis mutandis* for moral facts.

Response: Rather than relativizing moral and practical claims to a choice of temporal counterpart relation, the stage theorist should simply insist that there is a particular counterpart relation – the relation I've called R – such that what I ought (both practically and morally) to do now depends on what will be true of me relative to *that* counterpart relation. The purpose of a theory of personal identity, as the stage theorist conceives of it, is precisely

to figure out what this relation is. If I step into a teletransporter tomorrow, there's one sense in which I will survive (relative to the psychological-continuity counterpart relation) and a different sense in which I will cease to exist (relative to the bodily-continuity counterpart relation), according to the stage theorist. I can accept all this without being any the wiser about whether I ought to use the teletransporter. For that I need to know which relation is the one that *matters*; in other words, I need a theory of personal identity. Whatever the right answer to this question turns out to be, however, it *won't* commit us to a multitude of other objects coincident or co-located with every person.

5. Conclusion

I think the personite problem is a serious problem. Anyone who thinks that there are personites which are duplicates in this world of a person in some other world has to explain why we are justified in granting moral status to the latter but not the former. As Johnston (2017: 620) points out, the impetus over the centuries to expand our sphere of moral concern to individuals not previously included within that sphere has been rightly driven by the discovery of significant intrinsic similarities between those individuals and others to whom we had already granted moral status – if personites turned out to be intrinsically *identical* to some (albeit merely possible) thing with moral status, it's not clear on what rational grounds we could possibly resist the conclusion that they have moral status too. But this conclusion, if true, is disastrous – it's incompatible with even our most basic intuitions about how to lead a minimally ethical and well-ordered life.

In this paper, however, I have argued that although the personite problem is a problem for the worm theory, it does not affect the stage theory. And this, I think, makes a not inconsiderable difference to how we ought to regard the state of the dialectic between these two views. Considerations typically adduced in support of the stage theory over the worm theory include its ability to capture our intuitions about fission and fusion cases without a revisionary theory of counting (e.g. Hawley 2004: ch.5), its more unified solution to the paradoxes of coincidence

(Sider 2001: ch.5) and its better fit with the manifest image (Parsons 2015) – to that list, I would simply add that the stage theory is compatible with our basic ethical and prudential principles and the worm theory is not. So much the worse, then, for the worm theory.

Johnston mentions the stage theory only once in his two articles, at the very end of his (2017). In that passage, he appears to suggest that even if the stage theory avoids the personite problem, it would be wrong to reject the worm theory in favour of the stage theory on these grounds alone because there is a *modal* analogue of the personite problem that affects all theories of persistence equally. Presumably the view Johnston has in mind is one according to which, for every plurality of properties instantiated by an object *O* at a time *t*, there is something coincident with *O* at *t* that instantiates those and only those properties essentially. Coincident with Theresa May, for example, is an object which instantiates the property of being called ‘Theresa’ essentially, which would go out of existence were Theresa May ever to decide to change her name. Call this view *modal universalism*. If modal universalism were true, both the stage theory and the worm theory would indeed be committed to the existence of a multitude of things coincident with and intrinsically identical to me, which would cause just as much havoc for ethics as the personite problem. What isn’t clear, however, is why we wouldn’t be justified in rejecting modal universalism on the very same grounds.⁶ There is a modal analogue of the stage theory, of course, namely *modal* counterpart theory; perhaps a version of the argument presented here could also be used to defend that view over its rivals. In any case, the more limited conclusion of this paper – that the stage theory avoids the personite problem – remains unchallenged.⁷

References

⁶ Johnston insists that any alternative to modal universalism would be inconsistent with naturalism, but other than citing an unpublished manuscript on the topic he does not elaborate on this point.

⁷ Thanks to Mark Johnston, and to the audience at the 2018 Joint Session in Oxford for helpful feedback on earlier versions of this paper.

Hawley, K. 2004. *How Things Persist*. Oxford: Oxford University Press.

Johnston, M. 2016. Personites, maximality and ontological trash. *Philosophical Perspectives* 30: 198-228.

Johnston, M. 2017. The personite problem: should practical reason be tabled? *Noûs* 51: 617-44.

Lewis, D. 1986. *On the Plurality of Worlds*. Oxford: Blackwell.

Magidor, O. 2016. Endurantism vs. perdurantism?: a debate reconsidered. *Noûs* 50: 509-32.

Parsons, J. 2015. A phenomenological argument for stage theory. *Analysis* 75: 237-42.

Sider, T. 2001. *Four Dimensionalism: An Ontology of Persistence and Time*. Oxford: Oxford University Press.