Four-Dimensional Objects

PETER VAN INWAGEN

SYRACUSE UNIVERSITY

It is sometimes said that there are two theories of identity across time. First, there is "three-dimensionalism," according to which persisting objects are extended in the three spatial dimensions and have no other kind of extent and persist by "enduring through time" (whatever exactly that means). Secondly, there is "four-dimensionalism," according to which persisting objects are extended not only in the three spatial dimensions, but also in a fourth, temporal, dimension, and persist simply by being temporally extended.

In this paper, I shall argue that there are not two but three possible theories of identity across time, and I shall endorse one of them, a theory that may, as a first approximation, be identified with what I have called "three-dimensionalism." I shall present these three theories as theories about the ways in which our names for persisting objects are related to the occupants (or the alleged occupants) of certain regions of spacetime.

I

Let us begin by considering some object that persists or endures or exhibits identity across time. I will use Descartes as an example of such an object. Let us draw a spacetime diagram that represents Descartes's "career." In order to confer on this diagram maximum powers of accurate representation, let us pretend two things: (1) that the diagram is three-dimensional—made of wire, say, with the z-axis perpendicular to the page—, and (2) that Descartes was a "flatlander," that he had only two spatial dimensions.
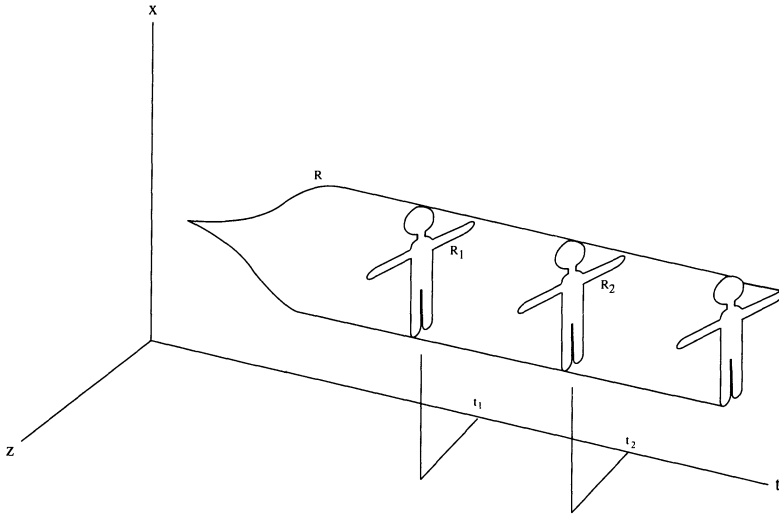
The outlined three-dimensional region in the diagram—or, since we are imagining that the "diagram" sticks out of the page and is made of wire, let us call it a model—represents a 2 + 1-dimensional region of spacetime called R. (We represent the dimensionality of regions of spacetime, and of objects that are extended in time as well as in space, by expressions of the form 'n + 1'. In such expressions, 'n' represents the number of spatial dimensions included in the region or exhibited by the object.) R is the region that some will say was occupied by the 2 + 1-dimensional Descartes; others will call it the union of the class of regions successively occupied by the always two-dimensional Descartes in the course of his career. R_1 and R_2 are subregions of R of zero temporal extent. Some will describe R_1 as the region occupied by the largest part of Descartes that is wholly confined to t_1; others will say that R_1 is the region that Descartes occupied at t_1. But, however R, R_1, and R_2 are to be described in terms of their relations to Descartes, it's clear which spacetime regions—that is, which sets of spacetime points—they *are*.

We may now present three theories about how the name 'Descartes' is semantically related to the occupants of R and of subregions of R like R_1 and R_2. (Two of these theories, the second and the third, are reflected in the disagreements I have noted about how to describe R and R_1.)

THEORY 1
If you say, "Descartes was hungry at t_1," you refer to, and ascribe hunger to, a two-dimensional object that occupies (fits exactly into)

R_1 and no other spacetime region. If you say, "Descartes was thirsty at t_2," you refer to a *distinct* two-dimensional object, one that occupies R_2, and ascribe thirst to *it*. Let us suppose that both of the descriptions 'the philosopher who was hungry at t_1' and 'the philosopher who was thirsty at t_2' denote something. It is evident that they cannot denote the *same* thing. It is therefore evident that the sentence 'the philosopher who was hungry at t_1 = the philosopher who was thirsty at t_2' cannot be true. Thus, if those sentences of ordinary English that *appear* to assert that one and the same person (building, river . . .) existed at two different times are ever to be true, what looks like the 'is' of identity in them must be interpreted as standing for some other relation than identity—call it gen-identity.

THEORY 2
When you use the name 'Descartes' you always refer to the 2 + 1-dimensional whole that occupies R. When you say "Descartes was hungry at t_1," you are referring to this whole and ascribing to it the property of having a t_1-part that is hungry. Thus, this sentence is exactly analogous to 'Water Street is narrow at the town line': in saying *that*, you refer to the whole of Water Street and ascribe to it the property of having a narrow town-line-part. What occupies R_1 is not what anyone, ever, even at t_1, refers to as 'Descartes'; it is, rather, a proper temporal part of the single referent that 'Descartes' always has.

THEORY 3
All of the regions like R_1 and R_2—instantaneous "slices" of R—are occupied by the very same two-dimensional object. When we say that Descartes was hungry at t_1, we are saying either (take your pick) that this object bore the relation *having* to the time-indexed property *hunger-at-t_1*, or else that it bore the time-indexed relation *having-at-t_1* to hunger.

The proponent of Theory 3, then, agrees with the proponent of Theory 2 that 'The philosopher who was hungry at t_1 = the philosopher who was thirsty at t_2' can be a genuine identity-sentence and be true; and he agrees with the proponent of Theory 1 that each of the terms of this sentence refers to a two-dimensional object—or, in the real world, a three-dimensional object. (But this second parallel should not be pressed too far. The "Oneist" holds that the terms of this sentence refer to objects that have non-zero extent in the spatial dimensions, but zero temporal extent: in *that* sense they are two-dimensional in our imaginary world, and three-

dimensional in the real world. The "Threeist", on the other hand, is probably not going to want to talk about temporal extent at all, not even temporal extent of zero measure. I shall presently return to this point.)

I am a proponent of Theory 3. In this paper, I can't hope to say even a fraction of what there is to be said about the questions raised by these three theories. I want to do just two things. First, to address some arguments for the conclusion that Theory 3 is incoherent, and, secondly, to present an argument for the conclusion that Theory 2 commits its adherents to a counterpart-theoretic analysis of modal statements about individuals. That hardly constitutes a refutation of Theory 2, of course, but, if true, it is an important truth; and it does seem that most philosophers, including, I suppose, many adherents of Theory 2, find counterpart theory rather unattractive. (I will not further discuss Theory 1, except in relation to one very special point. I doubt that anyone would prefer Theory 1 to Theory 2.)

<div align="center">II</div>

In this section, I shall reply to four arguments for the conclusion that Theory Three is incoherent. I shall also attempt to answer two pointed questions that my replies to these arguments are likely to raise.[1]

Argument A What exactly fills one region of spacetime cannot be what exactly fills another.

Reply Any plausibility that this assertion may have arises from an illegitimate analogy with the clearly true principle:

> What exactly fills one region of *space* at a given time cannot be what exactly fills a distinct region of *space* at that time.

This is valid for a space of any number of dimensions. Suppose spacetime is 9 + 1-dimensional, as in "superstring" theories. Then space is nine-dimensional and what occupies any, e.g., four-dimensional region of space at t is not what occupies any other four-dimensional region at t—much less any two- or seven-dimensional region. But the corresponding *spacetime* principle is wrong, or at least not self-evident, and would be wrong, or not self-evident, for any number of dimensions.

The spacetime principle may get an illusory boost from our three-dimensional physical model of a 2 + 1-dimensional spacetime. The two-dimensional region of space that represents R_1 in the physical

model, and the two-dimensional region of space that represents R_2 in the model, cannot, of course, be simultaneously occupied by the same two-dimensional physical object. But it no more follows that R_1 and R_2 must have different occupants than it follows from the fact that two photographs are in different places at the same time that they are not photographs of the same object. Our model occupies a three-dimensional region of space; one axis of the model has been arbitrarily assigned the task of representing the temporal dimension of a 2 + 1-dimensional spacetime. But this three-dimensional region of space is simply *not* a 2 + 1 dimensional region of spacetime, and the properties of a 2 + 1-dimensional region of spacetime can be read from the model only with caution. In my view, at least, any support that the physical model seems to give to the spacetime principle is an "artifact of the model." We could perhaps imagine a universe—call it Flatland—associated with a 2 + 1-dimensional spacetime, a universe whose spatial dimensions at different times coincided with those of appropriate cross-sections of the model. If the speed of light in Flatland were low enough, time-like intervals in the spacetime of Flatland might even be made to coincide in a non-arbitrary way with appropriate spatial intervals in the model. Nevertheless, the space the model occupies would not be a *duplicate* of the spacetime of Flatland, but only a representation of it.

Argument B (The "Twoist" speaks.) "Do you say that only a part of Descartes occupies R_1, or that all of him does? In the former case, you agree with me—in the latter, well it's just obvious that you haven't got all of him in there."

Reply I cannot yet answer this question because the appropriate senses of 'part of' and 'all of' have not yet been defined. I shall return to this question. For the nonce, I will say that my position is that *Descartes* occupies both R_1 and R_2, and that if you understand 'part of Descartes' and 'all of Descartes', then you understand 'Descartes'.

Argument C Theory 3 must employ either time-indexed properties or the three-term relation "x has F at t." But how are these properties, or this relation, to be understood? Take the case of the relation. We are familiar with the relation "x has F," the relation that holds between an object and its properties. If we are to understand the three-term relation, we must be able to define it using the two-term relation and other notions we understand. (We cannot simply take "x has F at t" as primitive, for that would leave the logical

connections between the two-term and the three-term relation unexplained.) The "Twoist" has such a definition:

x has F at t = df the t-part of x has F.

But the "Threeist" has no such definition. He must leave the relationship between *has-at-t* and *has* a mystery—and a wholly unnecessary mystery, at that. One might as well postulate a mysterious, inexplicable connection between "x has F" and "x has F at the place p." Just as it is obvious that 'The U.S. is densely populated in the Northeast' means 'The northeastern part of the U.S. is densely populated', it is obvious that 'The U.S. was sparsely populated in 1800' means "The 1800-part of the U.S. was sparsely populated'.

Reply One may say both that the relation "x has F at t" is primitive and that its connection with "x has F" is not inexplicable. One need only maintain that "x has F" is the defined or derived relation, and "x has F at t" the undefined or primitive relation. (Such cases are common enough. Consider, say, "x is a child of y" and "x is a child of y and z.") And I do maintain this. To say that Descartes had the property of being human is to say that he had that property at every time at which he existed. To say that he had the property of being a philosopher is to say that he had that property at every member of some important and salient class of moments—his adult life, say. I concede that "x has F" is primitive and "x has F at the place p" is derived (or, more exactly, that "x has F at t" is primitive and "x has F at t at p" is derived). But I see no reason why I should take the interaction of place and predication as a model for the interaction of time and predication. It may be that both space and time are abstractions from the concrete reality of spacetime. But they are *different* abstractions, and may be differently related to many things, including predication.

Argument D What occupies R_1—call it D_1—is clean-shaven. What occupies R_2—call it D_2—is bearded. Hence, D_1 is not identical with D_2.

Reply R_1 and R_2 are *indices*. Descartes is clean-shaven at R_1 and bearded at R_2. Let R_3 be a region of spacetime that was occupied by Mark Brown at some instant in 1973. I could point at Brown and say (correctly), "See that bearded man over there? He is clean-shaven at R_3."

Pointed Question 1 So "that man over there" occupies R_3, a region that fell within 1973. *When* does he occupy it?

Answer When is the proposition that Descartes was born on March 31st, 1596 true? Say what you like: that it's timelessly true, that the question is meaningless, that it's always true, that, strictly speaking, there is no *time* at which it's true. . . . and I'll obligingly adopt the corresponding answer to your question.

Pointed Question 2 So Descartes occupies both R_1 and R_2. What occupies R? And what properties does it have? Please describe them carefully.

Answer Well, it's not clear that I'm forced to say that *anything* occupies R. But let's assume that something does. It seems plausible to suppose that if something occupies R_1 and R_2 then, if anything occupies $R_1 \cup R_2$, it must be the mereological sum of what occupies R_1 and what occupies R_2. And it seems plausible to generalize this thesis: if something occupies the union of a class of regions of spacetime, and if each member of that class is occupied by something, then the thing that occupies the union must be the mereological sum of the things that individually occupy the members of the class.

Now the region R is the union of an infinite class of regions that includes R_1 and R_2 and indenumerably many other regions much like them. Each of these regions, *I* say, is occupied by, and only by, Descartes. It follows from this and our "plausible supposition" that it is *Descartes* that occupies R.

You ask me to describe carefully the properties of this object. An historian of early modern philosophy could do this better than I, but I can certainly tell you that it was human, that it was French, that it was educated by the Jesuits, that it wrote the *Meditations on First Philosophy*, that it believed that its essence was thinking, that it died in Sweden, and many things of a like nature.

Of course, the question is a little imprecise, since the occupant of R had different properties at different indices—it was, for example, hungry at R_1 and full at many other regions. If you insist on treating *R* as an index, and ask what properties the occupant of R had *at* R, it seems most reasonable to say: only those properties that it had at *all* the "momentary" indices like R_1 and R_2: *being human*, say, or *having been born in 1596*.

We may note that if Descartes occupies R as well as R_1 and R_2, this explains why the adherent of Theory 3 and the adherent of Theory 1 cannot mean quite the same thing by saying that the referent of, e.g., 'the philosopher who was hungry at t_1' is—in the real world and not in our simplified 2 + 1-dimensional world—a three-dimensional object. The "Oneist" means by a three-

dimensional object (at least in this context) one that has a greater-than-zero extent in each of the three spatial dimensions, and zero extent in the temporal dimension. But the "Threeist," if he takes the option we are now considering, believes that Descartes occupied R_1, which is of zero temporal extent, and *also* occupied R which has a temporal extent of fifty-four years—and, presumably, that he occupies regions having extents whose measures in years correspond to every real number between 0 and 54. Therefore, in his view, Descartes did not have a unique temporal extent. That is to say, he didn't have a temporal extent at all; the concept of a temporal extent does not apply to Descartes or to any other object that persists or endures or exhibits identity across time. Thus, in saying that the philosopher who was hungry at t_1 was a three-dimensional object, the "Threeist" means that he had a greater-than-zero extent in each of the three spatial dimensions—and that's all.

This completes my attempt to meet the most obvious arguments for the incoherency of Theory 3. I now turn to the promised argument for the conclusion that Theory 2 commits its adherents to a counterpart-theoretical understanding of modal statements about individuals.

<div align="center">III</div>

Theory 2 entails that persisting objects, objects like Descartes, are sums of *temporal parts*. That is, the "Twoist" holds that persisting objects are extended in time, and are sums of "briefer" temporally extended objects. Descartes, for example, extended from 1596 to 1650, and, for any connected sub-interval of that fifty-four year interval, that sub-interval was occupied by a temporal part of Descartes. (He may also have had discontinuous or "gappy" temporal parts, but, if so, we shall not need to consider them.)

Now it does not seem to be the case that Descartes had a temporal extent of fifty-four years essentially: his temporal extent might have been one year or fifty-five years or even a hundred years. But how will the Twoist understand this modal fact, given his thesis that Descartes is an aggregate of temporal parts? He will almost certainly not say *this*: If Descartes had had a different temporal extent from his actual temporal extent, he would have been composed of exactly the same temporal parts that composed him in actuality, but some or all of those parts would have had a different temporal extent from their actual temporal extent. For example, it is not likely that the Twoist will say that if Descartes had had a temporal extent of eighty-one years, he would have been composed of exactly the same temporal parts, each of which would have had a temporal

extent half again as great as its actual temporal extent. No, the Twoist will want to say that if a temporally extended object like Descartes has different temporal extents in different possible worlds, it must accomplish this feat by being the sum of different (although perhaps overlapping) sets of temporal parts in those worlds. And the Twoist will want to say this because he will want to say that temporal parts (i.e., objects that are temporal parts of something) have their temporal extents *essentially*. The Twoist will want to say that it would make no sense to say of the temporal part of Descartes that occupied the year 1620 that it might have had an extent of a year and a half: any object in another possible world that has a temporal extent of a year and a half is some other object than the object that in actuality is the 1620-part of Descartes. We may summarize this point by saying that the Twoist will want to maintain that temporal parts are "modally inductile" (and "modally incompressible" as well). And I am sure that the Twoist is right to want to say these things. If there are objects of the sort the Twoist calls temporal parts, then their temporal extents must belong to their essence.

But then the argument against Theory 2 is almost embarrassingly simple. If Theory 2 is correct, then Descartes is composed of temporal parts, and all temporal parts are modally inductile. But Descartes himself is one of his temporal parts—the largest one, the sum of all of them. But then Descartes is himself modally inductile, which means he could not have had a temporal extent greater than fifty-four years. But this is obviously false, and Theory 2 is therefore wrong.

We may also reach this conclusion by a slightly different route. If Theory 2 is correct, then there is an object, a temporal part of Descartes, that we may call his "first half." Now suppose that Descartes had been annihilated halfway through his actual span: then Descartes would have *been* the object that is in actuality his "first half." (At least I think so. In a possible world in which Descartes ceased to exist at the appropriate moment, Descartes would have existed—we have so stipulated—and so would the object that is, in actuality, his first half. At least I *think* it would have. How not? But if they both existed in such a world, what could the relation between them be but identity?) But if Descartes and a numerically distinct object could have been identical, then they conspire to violate the very well established modal principle that a thing and another thing could not have been a thing and itself.

There seems to me to be only one way for the Twoist to reply to these arguments. The Twoist must adopt a counterpart-theoretic

analysis of modal statements about individuals. And he must suppose that there are two different counterpart relations that figure in our modal statements about the object X that is both the person Descartes and the largest temporal part of Descartes: a *personal* counterpart relation and a *temporal-part* counterpart relation. According to this view of things, an object in some other world will count as a temporal-part counterpart of X only if it has the same temporal extent as X—anything that lacks this feature will be *ipso facto* insufficiently similar to X to be a counterpart of X under that counterpart relation. But an object in another world will count as a personal counterpart of X only if, like X, it is a maximal aggregate of temporal parts of persons. (That is, only if it is a temporal part of a person and its mereological union with any temporal part of a person that is not one of its own parts is not a temporal part of a person.) This device will allow us to say that X, which is both a temporal part and a person, could not have had a greater temporal extent *qua* temporal part and could have had a greater temporal extent *qua* person. That is: while every temporal-part counterpart of X has the same temporal extent as X, some personal counterparts of X have greater temporal extents than X. (As to the second argument: (i) counterpart theory allows world-mates to have a common counterpart in another world; (ii) this liberality is irrelevant in the present case, for if an object Y in another world is a maximal aggregate of temporal parts of persons that is an intrinsic duplicate of the first half of X, Y will not be a counterpart of *both* X and the first half of X under either counterpart relation.)

This reply to our two arguments is certainly satisfactory—provided that one is willing to accept counterpart theory. (It is important to realize that, as Stalnaker has pointed out, one can accept counterpart theory without accepting the modal ontology—David Lewis's ''extreme'' or ''genuine'' modal realism—that originally motivated it.[2]) I can see no other satisfactory reply to these arguments. I conclude that the proponents of Theory 2 are committed to a counterpart-theoretic analysis of modal statements about individuals.[3]

NOTES

[1]Three of the arguments—A, B, and D—and the pointed questions are taken from letters I have received from, and conversations I have had with, various philosophers. I am particularly grateful to David Armstrong, Mark Heller, Frances Howard, Michael Levin, David Lewis, and Michael Patton. Argument C is an adaptation of some points that have been made by David Lewis. See his discussion of ''the problem of temporary intrinsics'' in *On the Plurality of Worlds* (Oxford: Basil Blackwell, 1986), pp. 202-204, and 210.

[2]Robert Stalnaker, ''Counterparts and Identity,'' *Midwest Studies in Philosophy* 11 (1986), pp. 121-140.