Time, Fission, and Personal Identity

John Perry

Stanford University

Abstract

I argue that the account I gave of Derek Parfit's dividing self-case in "Can the Self Divide?" does not depend on a dubious four-dimensionalist metaphysics as claimed by Eric Olson (2006). I explain my metaphysics of time, and then re-describe my solution in "Can the Self Divide?" comparing it to treatments of the dividing selves by Parfit and Lewis.

Identity, Unity relation, Four-dimensionalism, Thoughts, Circumstances, Events, Person-stages.

1. Introduction

Derek Parfit was a wonderful philosopher and an exemplary human being. The only opportunity I had to interact with him personally for any length of time was at a symposium at the University of North Carolina in the early 1970's. Parfit, David Lewis, and I gave talks on personal identity, focused on the problems Parfit had raised in his watershed 1971 paper, "Personal Identity".[1] It was a thrill to meet and talk with Parfit, and I regret that our paths did not cross more often. David Lewis was already a good friend by then. I can't remember a symposium I enjoyed more, or from which I learned so much.

Parfit's key examples were cases of fissioning or dividing selves. One person becomes two, both of whom have first-person memories and meet other criteria for being the same person as the original. But they can't both be the original person, since are not each other. They go their different ways with no special access to each others' minds. That shows, Parfit argued, that identity isn't really the important relation. Most of our discussion concerned such cases. Lewis and I thought we could salvage a little more importance for identity than Parfit granted, if we could accommodate the language of identity to the cases of division. In my article "Can the Self Divide?"[2] I had concluded that in a case of fission we had to recognize three persons. In a two-dimensional representation of

[1] Parfit 1971.

[2] Perry 1972. Reprinted in Perry 2002. I had not seen Parfit's "Personal Identity" when this was written. Parfit's essay was being accepted by the *Philosophical Review* at about the time my essay was being rejected.

Argumenta 5, 1 (2019): 11-20
ISSN 2465-2334

DOI 10.14275/2465-2334/20187.PER

the case, one would have a "Y" shaped lifetime, the others would have lifetimes corresponding to the branches of the "Y". Lewis didn't think I had things quite right: there were only two persons involved, corresponding to each of the branches. Parfit found it difficult share our worries about exactly what to say about identity in fission cases, so long as we were clear about the relations that really mattered. But there seemed to be a basic agreement on the following:

> Identity, a necessarily transitive relation, needs to be distinguished from the temporal unity relation for persons (my terminology). The latter obtains among events that are part of person's life. As Parfit examples of fission and fusion showed, it is not necessarily transitive.
>
> Locke and others who argued in effect that first-person memory was at least an important component of the unity relation were on the right track.
>
> Parfit was right that the aspects of the unity relation he called contiguity and connection were important and important to distinguish, whether or not he was right that identity itself was unimportant.

In the ensuing years, it has been argued that the position I held in this discussion requires one to hold one or more controversial metaphysical views. Eric Olson says in his review of my collection, *Identity, Personal Identity and the Self*, with "Can the Self Divide?" particularly in mind:

> Perry says a great many things about identity over time that make perfect sense against the background of a temporal-parts ontology but are outrageous otherwise. This matters because the ontology of temporal parts—better known as 'four-dimensionalism'—is a contentious metaphysical claim.[3]

In this essay I sketch a view of time that, if outrageous, is not so in quite the way Olson seems to have in mind. Then I offer what I hope is a somewhat clearer and more plausible version of the approach in "Can the Self Divide?".

2. A Framework

In this section I explain my framework for thinking about time, inspired by Alfred North Whitehead and Gottlob Frege, the subjects of my undergraduate and graduate theses. But I don't accept all of what either of these great philosophers had to say. It is a version of or at least a descendent of the framework of situations and states of affairs in *Situations and Attitudes*.[4]

I'll start with what Frege tells us in "The Thought".[5] There are three realms, two causal, the physical and the mental, plus a Third Realm of abstract objects, including what he calls "Thoughts". I'll capitalize the word when using it in Frege's sense. I'll call my version of the first two realms "reality" and my version of the third "possibility".

Frege's Thoughts are abstract objects that do not change in their intrinsic nature. They don't cause anything, and aren't effected by anything. They don't involve objects like you or I. They delineate patterns of the instantiation and co-instantiation of properties and relations. According to Frege Thoughts *determine*

[3] Olson 2006: 437.
[4] Barwise and Perry 1983.
[5] Frege 1956.

truth-values; a sentence refers to the truth-value determined by the Thought it expresses, either necessarily, or in conjunction with what happens in reality. I'll just say that Thoughts can be true or false, treating truth and falsity as properties, a practice Frege recommended against.

I think we need to recognize that Thoughts do change in two extrinsic ways, due to what happens in reality. First, physical and mental events occur at various times, completing the conditions necessary to make a contingent Thought true or false. The word "completing" is important. Dwight Clark's 1982 end zone catch of Joe Montana's pass made it the case that the 49-ers beat Dallas and became NFC champions—but only because 49-ers were less than a touchdown back with seconds left. The catch completed the job.

Second, minds grasp Thoughts at various times, coming to believe them, doubt them, reject them, or whatever. On Frege's view, that is the only practical importance of Thoughts. I sort of agree. Thoughts help us keep track of the contents of such grasping episodes: the truth-conditions of utterances and beliefs, the satisfaction conditions of desires, and so on. The same goes for the additional abstract objects I introduce below. Perhaps we should think of them as theoretical posits, the theory being that reality makes a certain amount of sense. My view may not motivate the reverence for the third realm that Frege would find appropriate. I would like to have an account of the realm of possibility that made it more a product human attempts to understand cognition and treat it systematically than he might approve of. But Frege doesn't tell us much about his views on such issues, and I don't have such an account.

On my interpretation, Frege's Thoughts are structures of properties. A key concept in his *Begriffsschrift*[6] was *falling under*; it provided the framework for his inventions of quantification and first and second order logic. Objects fall under first-order properties and relations. So "Caesar conquered Gaul" tells us that Caesar and Gaul fall under the relation of conquering. The sentence refers to (*bedeutet*) a circumstance (*Umstand*) which is its conceptual content. Since Caesar did conquer Gaul, this circumstance is a fact. "Some Roman conquered some country" tells us that the properties of being a Roman and being a country fall under the higher order relation of there being someone with the first that conquered something with the second. We could call what this sentence refers to a "higher-order circumstance", but I will call it a "property-structure" and reserve "circumstance" for first-order circumstances. I'll use "property" for "properties and relations" unless indicated otherwise.

The combination of views Frege held, especially that every function must be defined on every argument, seem to lead to contradiction. But the basic ideas can be incorporated into contemporary theories of higher-order logic, if we give up that view and adopt some version of type theory.

This is my interpretation of the *Begriffsschrift*. Frege seems to find the concept of a circumstance intuitive for what he wants to say, but he doesn't treat it as a technical term. Frege's problems with identity eventually led him to give up circumstances as part of his semantics, and replace conceptual content with sense and reference. Frege wanted an account of content that captured "cognitive value" or "cognitive significance". But "Hesperus is moonless", "Phosphorus is moonless", and "Venus is moonless" differ in cognitive significance while referring to the same circumstance and having the same conceptual content. In

[6] Frege 1979.

the *Begriffsschrift*, Frege tried to solve the problem by removing "=" from his language in favor of "≡" which refers to the relation between names of having the same referent. (Contrary to his inaccurate description of his earlier view in "On Sense and Reference", he did not take identity to be a relation between expressions.) But eventually, by the time he wrote "Function and Concept",[7] he realized that the problem was with circumstances, and not with "=", as is shown by the example above, which does not contain "=". In the theory of sense and reference, circumstances disappear, and "=" returns. The Thoughts expressed by the problematic sentences do not contain Venus, but different senses that determine Venus as reference.

I interpret the Thoughts of the theory of sense and reference as basically the property structures of the *Begriffsschrift*, invested with additional duties as Frege's ideas continued to develop. I prefer Thoughts, construed this way, to contemporary theories of propositions in terms of possible worlds. If a world is a *totality* of facts, I doubt that we have even one such world, for reasons explained in *Situations and Attitudes.*

I think we need two additional sorts of abstract objects in order to adequately provide for the contents of cognitions.

I agree with Frege that circumstances cannot encode truth-conditions in a way that captures the cognitive value or cognitive significance of sentences. But I think we need circumstances to deal with the way ordinary humans store and communicate information, in what Frege calls "imperfect" languages. He gives the example of a conversation about Aristotle, where the participants attach different senses to the name "Aristotle". He says that such divergence is "tolerable" in using imperfect languages, but doesn't tell us much more. It seems that what makes it tolerable is that although the participants express different Thoughts with, say, "Aristotle liked syllogisms", they refer the same circumstance. Our methods of information transmission, from casual conversations about Aristotle to radios, television, and the internet are devised to insure sameness of circumstance. There may be constraints on how different the Thoughts can be. If the announcer had said "The only 49-er to score a touchdown in the Gator Bowl just scored another one", fans would have complained. But sameness of Thought is seldom important for practical purposes.

If he had second thoughts about excluding circumstances from his philosophy, where would Frege have put them? Since they contain objects, they wouldn't be allowed in the Realm of Thoughts. But I won't worry about that issue of counterfactual interpretation. I put circumstances in the realm of possibility. But whereas Thoughts are in Frege's third realm independently of time, it seems that circumstances *become* possibilities only when the objects involved in them becomes available. So my realm of possibility is less eternal than Frege's.

Just as some Thoughts are made true by events, some circumstances are made *factual* by events. Circumstances are basically what are now called "singular propositions" or perhaps a subset of them. I prefer to use "circumstance" because I attach some importance to the distinction between being factual and being true, but that won't matter here.

We also need a third kind of abstract object, which I call an "episode relative role", or "role" for short. These are partial functions that take us from an agent and time of an episode of perceiving, thinking, speaking, or acting to a

[7] Frege 1997.

circumstance.[8] Consider the visual state of any chicken, when it sees a kernel of corn in front of it. An occurrence of the state in a particular chicken at a particular time will be *veridical* only if there is a kernel of corn in front of *that* chicken at *that* time. Not any perceptual state that has those conditions of veridicality will do. If a chicken sees herself in a mirror with a kernel of corn in front of her, she won't peck. She has to have the information that there is a kernel in front of her *in a certain way*. Instances of the perceptual state in different chickens at different times will represent different circumstances. But they will all have the same causal role: causing the chicken in the state to peck at the time it is in the state (or very soon thereafter). Chicken theorists need both circumstances and roles to explain what is going on.

Presidential theorists also need them. When an Iraqui journalist threw a shoe at George W. Bush, he ducked. We say this was just a reaction, but it was quite rational. He gained information perceptually, that a shoe was headed toward him. He had a desire not to get hit in the face with a shoe. He took an action that, in the circumstances he perceived, achieved the desired result. This was primitive self-knowledge at work. He didn't need to think "There is a shoe coming at *my* head *now*", much less "There is a shoe coming at George W. Bush at 2 p.m. on December 14 2008". If he had, he wouldn't have ducked in time. As sophisticated as we humans are, or think we are, we never outgrow our need for primitive self-knowledge. Some philosophers think in terms of a representation of reality that abstracts from primitive self-knowledge, and then complain that they cannot find the self and the now in the representation, and draw metaphysical conclusions. I call this "the fallacy of abstraction".

Once an event has made a Thought true, or a circumstance factual, or a perception veridical, no later event can make them false, mere possibilities, or illusory. Consider the sentence, "George Bush was hit by a shoe on December 14 2008". Assume we have a sense for "George Bush" and so a Thought T. It seems that T was made false on December 14, after Bush ducked the last of the shoes thrown at him and retired to his hotel room where no one could throw shoes at him for the rest of the day. So it seems that on December 13, T was not true, and then it was made false, so by December 14 it was false and has remained so ever since. This suggests the need for a three-valued logic. Frege didn't offer one. If we define "be true" as "was, is, or will be made true" and analogously with "be false" we can stick with a two-valued logic for *being true* and *being false*.[9]

So much for abstract objects. What about the events that make some Thoughts true and some circumstances factual—what Whitehead calls "actual occasions" (more or less)?[10] I don't have anything very complete or sophisticated to offer. We see things happening. I'll just call them "happenings". Happenings make circumstances factual. We think of many happenings as the instantiations of circumstances: objects having properties and standing in relations in locations at times. I think my pipe is on the counter in the kitchen. The facticity-conditions of my thought are encoded by a circumstance consisting of my pipe, the kitchen counter, the relation of being on, and the present time. I go to the kitchen and see my pipe on the counter. I am not seeing a circumstance. I am

[8] Roles were inspired by David Kaplan's *characters*. See Kaplan 1979.
[9] See Perry 2006.
[10] Whitehead 1929.

seeing something going on in reality, a *happening*—admittedly not a very exciting one—that (incrementally) makes the circumstance a fact.

Like many expressions in English, "event" is used for both real things and possibilities. I have planned many events, which never happened. I have seemed to remember many events which did not happen. These "events" are circumstances that never became facts and never will. But we also say, after a party, "That was a fun event", or "I'm glad that event is over, it really tired me out". Here we seem to be referring to happenings in reality that can be seen and can tire one out. I will use the term "event" in the first way, for circumstances. When I say an event *happened*, I mean *happenings* made the circumstance factual.

Events happen at times. Once they have happened we say things like "that event is now in the past". It's better to just use that past tense. "That event happened". It's not now happening in some special place called the "past". Similarly, to say an event is "in the future" is to say it has not happened yet, but it will; that is, there will be happenings that make it factual.

There is, of course, an important difference between what has happened and what will happen. Once a circumstance has been made factual, or a Thought has been made true, it will never be made non-factual or false. But events that have not happened are not yet facts. Even if the issue is in some way settled, by laws of nature and past events, say, until the event happens it hasn't happened. It's not off happening in the future waiting to appear.

I once said I accepted the "growing block" theory of the past, as C. D. Broad once did and Michael Tooley does.[11] I'm not sure whether I have given up the growing block view, re-interpreted it, or merely finally understood it. We shouldn't think of the block as a storage bin of old events, getting more crowded with time, but as the series of large and larger subset of Thoughts that have been made true and of circumstances that have been made factual.

I think that the happenings I see happen involve smaller and smaller happenings I don't see. Perhaps there are basic happenings, or perhaps there is no end to the sequence of more and more basic happening. Perhaps we should think in terms of tropes instead of in terms of objects having properties. I don't have settled opinions, except in the sense that an opinion is settled when you don't plan think about the topic again for a while.

So that's my metaphysics of time, plus the caveats that I wish I understood Thoughts and other abstract objects better than I do, and I wish I understood events better than I do. I'm not presenting it as a comprehensive metaphysics of time, just as enough metaphysics to make my approach to the problems of personal identity seem sensible to some sensible philosophers.

3. The Case of the Dividing Self

"Can the Self Divide?" discussed a case of fission. Jones, Smith and Brown go into the hospital for "brain rejuvenations". Through a series of mishaps the bodies of Brown and Smith end up with duplicates of Jones' brain; Jones' body gives out on the operating table. How many persons should a memory-theorist suppose we have in this scenario? Call the survivors "Smith-Jones" and "Brown-Jones". Given my current views about time, it will simplify discussion

[11] See Perry 2006.

if we suppose that this all happened in the past, so all the events we are discussing have all happened.

I call the relation that holds between stages of the same person the "temporal unity relation for persons". I call Locke-inspired analyses of the temporal unity relation "mentalist", and I'll use "M" for the most plausible mentalist candidate. My intuitions were, and are, that some mentalist analysis of the temporal unity relation for persons gives the correct result. I also think that the mental is part of the physical, but I won't worry about that here. Jones was to do all of the things Smith-Jones and Brown-Jones went on to do after the operation; Smith-Jones and Brown-Jones did all of the things Jones did before the operation. But Smith-Jones and Brown-Jones do not do the same things after the operation. (I am using "do" as short for something like "do, experience, have happen to them, etc.") So Smith-Jones and Brown-Jones were not the same person. Thus, if we take Jones to be the same person as Smith-Jones and also the same person as Brown-Jones we have a problem. Does this mean that we need to abandon the mentalist analysis of the unity-relation? I argued that is does not.

A sequence of person-stages is a *branch* if and only if all the members of the set have *M* to one another, and no stage that has *M* to all the members of the sequence is not a member. A set of person-stages is *a lifetime*: there is some member in the set such that all and only members of the set have *M* to that person-stage. If we draw a two-dimensional diagram, the lifetime will be a Y-shaped structure with two branches.

One strategy is to suppose that we speak the *stage-language*. We speak loosely when we say the same person occurs at one place at one time and then at another place at another time. We simply have *M-related* stages. If we keep this in mind, fission cases don't present a problem.

Another strategy is to suppose that we speak the *branch-language*. The events in the life of a single person form a branch. This is more or less the view Lewis's defended.

A third strategy is to suppose that we speak the *lifetime-language*. The events in the life of a single person form a lifetime. This is my view. Or, more cautiously, if we add the lifetime language to English, we can take care of fission and fusion cases in a way that accounts for our intuitions and leaves English unmolested for the normal cases.

I'll say that any person-stage *determines* a lifetime, the set of all stages to which it has M. A lifetime is *determinable* at *t*, if a person stages that determine that lifetime occurs at *t*. A person-stage *identifies* a lifetime relative to a time, which need not be a time at which the stage occurs. A person-stage identifies a lifetime at *t*, if it is the unique lifetime determinable at *t* that contains the stage. Jones pre-operative stage determines the Y-shaped lifetime. It doesn't identify any lifetime relative to a time after the operation, since there is no unique lifetime determinable then contains it. A post-operative stage of Smith-Jones (or Brown-Jones) together with a time before the operation identifies the Y-shaped lifetime. In normal cases the lifetime a stage determines will also be the one it identifies relative to a time.

In the lifetime language, we distinguish between the *primary* and *secondary* referents of a name. Assume a name N is assigned to a person at a time, with reference to a person-stage *s* occurring at that time. The lifetime determined by that stage is the primary referent of N. The secondary referent of N at time *t* is the lifetime identified by *s* at *t*. If *s* does not identify a lifetime at *t*—if there is no

unique person-stage at *t* with *M* to *s*—then *N* has no secondary referent at *t,* and *N* is *improper* at *t*. Thus, in the ordinary case, the secondary referents of *N* and its primary referent will be one. But in unusual cases, they will not.

Consider our troublesome argument:

(1) Smith-Jones is the same person as Jones.
(2) Brown-Jones is the same person as Jones.
(3) Smith-Jones is the same person as Brown-Jones.

In (1) "Smith-Jones" is improper if uttered before the operation, and "Jones" is improper if used after the operation, and similarly for (2). (1) and (2) do not express propositions, and can't be used as premises for (3), even after the operation when (4) expresses a false proposition. So the argument (1)-(3) is blocked.

Here I am assuming that in sentences like (1), (2), and (3) the time with respect to which the names are proper or improper is the time of utterance. But if we embed the sentences after a temporal operator, the operator provides a second time relative to which a name can qualify as proper. If we say (1') after the operation, "Jones" will be proper relative to the time before the operation, "Smith-Jones" will be proper relative to the time of utterance and we will say something true, and similarly for (2').

(1') Before the operation, Smith-Jones was Jones.
(2') Before the operation, Brown-Jones was Jones.

So the argument (1)-(3) is blocked.

I think the account needs to be improved in two ways. First, I said I was assuming for most of the discussion that this all happened far in the past, with no participants still living. But then I use their names which are all now improper. Please assume that I am implicitly supplying a suitable temporal operator.

Second, so far we have only one relation between persons, personal identity, analyzed in terms of the M-relation. I think we need a second one. A natural thing to say is that Jones *became* Smith-Jones and Brown-Jones. I'll take *becomes* to be a relation between persons *a* and *b* that obtains if the lifetime of *a* includes the events in the lifetime of *b*. Most of us simply keep becoming ourselves, and are identical with the person we become. But Jones became Smith-Jones and Jones became Brown-Jones. A person *b came from* a person *a* if *a* became *b*. I'll also say that after the operation Jones *exists as* Brown-Jones and Smith-Jones, and before the operation they existed as Jones. And I'll say that a person *exists at a time* only when they exist as themselves at that time.

Lewis said:

> How many persons were involved in an episode of fission long ago? I say: two. Perry says: three... Isn't two the correct answer? (Lewis 1976: 39).

There was only one person in Jones' bed before the operation, Jones. Before the operation, the events that constituted fission had not occurred. Smith-Jones and Brown-Jones were no more than possibilities. And the operation could have been even more fouled up, so that Jones could have become three or four or a dozen persons. So, sitting there before the operation, we should definitely say that there is just one person in the hospital bed. The fact that *becomes* two other persons, a very intimate relation that usually amounts to identity, doesn't mean the persons he becomes are in bed with him.

I believe that any two events can be compared in terms of three spatial and one temporal dimension. Consider the event of my leaving Stanford at five p.m. and my arriving home at 5:30 p.m. The latter event occurred two miles east, one mile north, ninety feet below, and thirty minutes later than the former. So we have four dimensions. I accept that there are four dimensions. But I don't think that qualifies me to be a *four-dimensionalist* in one way that term is now used in philosophy. Such a four-dimensionalist thinks atemporally, so that in some way that past, present and future events are "equally real". Olson may be right that my view will seem more plausible to someone who is a four-dimensionalist and thinks "atemporally". And from that perspective, perhaps Lewis's view is more plausible. To draw a diagram like one he has in his article, with a Y shaped structure, we only have to draw two branch shaped lines, not three. I don't think that kind of four-dimensionalism is the perspective I had in 1972. But, in Parfit's terminology, my current stage is only connected with my 1972 stage not contiguous with it. But, in any case, my current stage wants to say that, before the operation, only one of the three people in our scenario existed.

In the original article I said that English was the lifetime language. That was a bit of an overstatement. I think that, for the purposes of discussing my fission cases, we can augment English with the lifetime language, without changing the truth-values or truth-conditions of any English sentences used to describe non-fission cases. Certain inferences that are pretty reliable in the world as it is won't be valid. For example, we legitimately infer that if a person did something at an earlier time, they existed at the time they did it. But we can explain why that inference works when we don't have fission.

I can't say that the reception over the past forty-five years demonstrates that my approach captures most philosopher's intuitions better than the alternatives.

But I do think the approach succeeds in showing we can describe fission cases coherently without giving up falling into contradiction, giving up on identity, or abandoning the analysis of personal identity in terms of a Locke-inspired unity relation. And it does not seem that it requires an appeal to any appeal to four-dimensionalism that amounts to more than common sense. And it does not require us to overlook Parfit's important and fundamental insights.

References

Barwise, J. and Perry, J. 1983, *Situations and Attitudes*, Cambridge (MA): The MIT Press. (Reprinted with a new introduction by CSLI Publications, 1999.)

Frege, G. 1879, *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert.

Frege, G. 1956, "The Thought: A Logical Inquiry", *Mind*, 65, 259, 289-311.

Frege, G. 1997, "Function and Concept", in M. Beaney (ed.), *The Frege Reader*, Hoboken: Wiley-Blackwell, 130-48.

Kaplan, D. 1979, "On the Logic of Demonstratives", *The Journal of Philosophical Logic*, 8, 81-98.

Lewis, D. 1976, "Survival and Identity", in Rorty 1976, 17-40.

Olson, E. 2006, "Review of J. Perry, Identity, Personal Identity, and the Self", *European Journal of Philosophy*, 14, 434-37.

Parfit, D. 1971, "Personal Identity", *Philosophical Review*, 80, 3-27.

Perry, J. 1972, "Can the Self Divide?", *Journal of Philosophy*, 69, 463-88.

Perry, J. 2002, *Identity, Personal Identity and the Self*, Indianapolis: Hackett.

Perry, J. 2006, "How Real are Future Events?", in F. Stadler and M. Stöltzner (eds.), *Time and History, Proceedings of the 28th International Wittgenstein Symposium, Kirchberg am Wechsel, Austria, 2005,* Frankfurt: Ontos, 13-30.

Rorty, A. (ed.) 1976, *The Identities of Persons*, Berkeley and Los Angeles: University of California Press.

Whitehead, A.N. 1929, *Process and Reality. An Essay in Cosmology*, New York: Macmillan; Cambridge: Cambridge University Press.