

Identity: this time it's personal

Stephen Kearns, Florida State University

Abstract: The view that it is possible for someone to think at a time without existing at that time is not only perfectly coherent, but in harmony with an attractive externalist view of the mental.

Furthermore, it offers plausible solutions to various puzzles of personal identity.

In what follows, I argue that it is possible for a person to think at a time without existing at that time. I thus reject the validity of what we might term the *synchronic cogito*—I think now, therefore I exist now.

1. Stop staring incredulously at me like that

Before I present my case, I shall address two sources of the following natural reaction: *of course it's not possible for someone to think at a time without existing at that time*. First, one might claim that it is impossible for something to be *any* way at a certain time without existing at that time. But this is false:

Some dinosaurs are fossilised now.

Vincent van Gogh is much more appreciated now than when he was alive.

Thanks to his discovering penicillin, Alexander Fleming is still saving lives.

Marie Curie influences today's chemistry.

Some dead stars illuminate tonight's sky.

These claims are all true and concern things that no longer exist (if one takes Timothy Williamson's line that all these things do still exist, albeit non-concretely (see his 2015), simply

read my use of “exists” as “concretely exists”—my main claim thus becomes that it is possible for someone to think at a time without *concretely* existing at that time). At the very least, the claim that it’s possible to think without at the same existing is no more *logically* problematic than these examples.

Second, one might claim that I have not learnt the lesson that Descartes taught us. That I exist is something I cannot (properly) doubt (Descartes 1993). This is because, roughly, it is necessary that if I doubt that I exist, then I exist. After all, if I did not exist, I wouldn’t be around to doubt my own existence. Doubting one’s own existence is thus self-defeating. *Cogito ergo sum.*

I agree with this reasoning as long as we are careful with times. The only thesis and inference I deny are the following:

It is impossible to think at a time without existing at that very time. (*I think at a certain time, therefore I exist at this very time.*)

I may consistently maintain the following pairs of theses and inferences:

It is impossible to think at a time without existing (at some time or other, or atemporally). (*I think at a time, therefore I exist (at some time or other, or atemporally).*)

It is impossible to think at a time without existing at a time. (*I think at a time, therefore I exist at a time.*)

It is impossible to think at a time without thinking and existing at a time. (*I think at a time, therefore I think and exist at a time.*)

It is impossible to think at a time without something's thinking and existing at that time.

(I think at a time, therefore something thinks and exists at that time.)

While it may be possible to establish the indubitability of my own existence via Cartesian-style reasoning (though Liebowitz, *forthcoming*, for an argument that it cannot even do this), such reasoning cannot establish the *synchronic cogito* in particular.

2. Externalizing the mental

Still, you might say, isn't obvious that thinking is different from the properties mentioned in the above examples (about Marie Curie, dinosaurs, etc.)? The above sentences are true because they describe how past things *relate* to objects and events at later times. They do not describe *how those past things intrinsically are* at these later times (while there have been many attempts to analyze the intrinsic/extrinsic distinction, probably the most prominent being Langton and Lewis 1998, the thrust of my view is compatible with many such understandings—going forward, I rely merely on the intuitive thought that something is some way intrinsically at a time just in case its being that way at that time does not constitutively depend on its relations to things at other times). When we think at a time, however, we are some way intrinsically at that time. And nothing can be some way intrinsically at a time without also existing at that time. Thus, when we think at a time, we exist at that time.

I will not dispute the claim that being some way intrinsically at a time requires existing at that time, but I reject the thesis that, necessarily, a person thinks at a time only if she is some way intrinsically at that time. That is, I claim that someone's thinking at a time may be due to how she *relates* to what is happening at that time, rather than how she is intrinsically at that time. Such an approach is in harmony with the various ways the mental is argued to be extrinsic to us.

Here is a sample of the existing ways in which the mental has been argued to be at least partly extrinsic:

Natural Kind Content Externalism: What we believe, desire, intend, etc. is in many cases partly constituted by the natural environment in which we're embedded. Our natural kind terms and concepts pick out natural kinds around us, which then figure in our beliefs (desires, etc.). Had there been other natural kinds around us, and even if we were intrinsically just the same, our natural kind terms and concepts would pick out different kinds, and we would thus have had different beliefs, desires, etc. This point is most famously argued in Putnam 1975.

Linguistic Community Content Externalism: What we believe, desire, intend, etc. is in many cases partly constituted by the linguistic community in which we're embedded. We are deferential in the use of some of our terms and concepts. That is, we use some terms and concepts to pick out whatever our linguistic community (or a subset of this community, such as certain experts) picks out with them. For instance, my use of the term "electron", and the beliefs I form about electrons, are parasitic on the behaviour and beliefs of my linguistic community. Had I been in a different such community, one which meant something different by "electron", I would have had different beliefs, even if I would have been intrinsically identical to how I actually am. This point is argued in Burge 1979.

Swamp Man: That we have beliefs, desires, intentions, etc. at all depends on the historical causal connections we have with our environments and communities. If a swamp were to be struck with lightning and an atom-for-atom duplicate of me (as I am currently) were

thus created, this swamp man would fail to have a mental life (at least initially), precisely because it would not yet be hooked-up to the world appropriately. This view is defended in Davidson 1987.

Knowledge as a Mental State: Knowledge cannot be factored into an intrinsic mental element (such as belief, if this is intrinsic), and an extrinsic non-mental element (like truth). Rather, knowledge is an unanalyzable *sui generis* factive mental state. That it is factive means that the mental is in part extrinsic—what I *know* about the external world depends on what is *true* about the external world. Thus, how I am mentally is in part constituted by how things are externally. This view is defended in Williamson 2000.

Extended Minds: Certain processes that occur outside my head can still count as partly constituting my mentality if they play the right role and are related in the appropriate ways to my other mental processes. For instance, suppose I often jot down thoughts and descriptions of events in a notebook which I then frequently consult to aid in my recall of such thoughts and events. This notebook plays a similar enough role to cognitive processes inside my head to count as a true cognitive process. The notebook and its uses are part of my extended mind. The mind thus extends beyond my bodily boundaries. See Clark and Chalmers 1998 for a defence of this view.

Adaptation Solutions to Deviant Causal Chains: I intentionally open the door if my opening the door is caused non-deviantly by my intentions. A cause is non-deviant just in case it is the type of causal process adapted (by evolution) or designed (by an agent) for that purpose. Thus, what we (intentionally) do is partly dependent on our evolutionary or design histories. For this type of view see Mele 2003, p53-63.

The Biological Purpose Response to Kripkenstein's Rule-Following Paradox: What we mean by our terms is not reducible to our finite use of these terms (Kripke 1982). Rather, what we mean is a matter of what rules we have a (derived) biological purpose to follow. What we have a biological purpose to follow is a matter, at least in part, of the evolutionary history of our species. Meaning is, then, partly constituted by our evolutionary history. See Millikan 1990 for this view.

Soft-Line Solutions to Manipulation Cases: Whether agents make free decisions or not depends on their history. If, in this history (exclusively at times at which the agents do not exist), powerful beings manipulated or designed these agents to live certain lives, and this manipulation/design ensured that the agents would indeed live these lives, then the agents do not make their choices freely. Agents intrinsically identical to the first agents, but who lack this manipulation/design in their history, can indeed make free choices. Thus, free will is an extrinsic matter. See, for example, Waller 2012, Herdova 2021.

Disjunctivism about Perception: When we perceive the external world, external objects constitute the contents of our perception. Hallucinations, though in some way perceptual, are radically different in nature from veridical perceptions, as they do not involve the external world (at least, constitutively). See Hinton 1967.

Post-Mortem Action: A person, S, shoots another, T. Before T dies of his wounds, S dies of a heart attack. When did S kill T? If we say S killed T before S died, then for some time after S killed T, T is still alive. Perhaps we should say that S completes her killing of T only when T dies. In which case, S completes her killing of T after S dies. One may complete actions after one no longer exists. Consider also cases of action-by-proxy. I can

inform someone of some news by passing on a message via an intermediary. If in the meantime, I cease to exist, I continue to act via the actions of this intermediary.

Survival without Identity: It is possible to survive one's own non-existence as long as one is appropriately psychologically connected to future people (one's psychological descendants). What it is rational to fear, anticipate, look forward to, etc. can depend on the psychological connections one has to distinct future people. If one lacks such psychological connections but is otherwise intrinsically identical to one who has them, then one will not survive the end of one's existence. See Parfit 1971, 1984.

Semantic Sovereignty: What we think and what our words mean not only fails to supervene on our intrinsic properties, but also on the physical properties of the entire universe. Indeed, what we think, and what we mean by our words, is in part determined by the presence or absence of possible non-physical beings with various semantic properties. See Kearns and Magidor 2012 for this view.

This list of ways in which the mental may be partly external is far from exhaustive (other possible sources of such ways include considerations from social epistemology, embodied cognition, phenomenology, the philosophy of religion, and others). Even if some of the above suggestions are incorrect, perhaps similar ideas *are* correct (for example, it may be that any viable solution to Kripkenstein's paradox, or the problem of deviant causal chains, must be externalist). The point is that there is a strong case that how we are mentally is to a significant degree a matter of how we are related to things outside us. Of course, each of the theses mentioned above spell out this idea in very different ways—the externalisms of Putnam and Burge emphasise how the *contents* of our attitudes depends on the situations in which we're

embedded, while the extended mind hypothesis of Clark and Chalmers claims that the *vehicles* of such contents may exist outside our skulls. Still, the pictures each of these views paint is of our mentalities as inextricably linked with the external world.

My thesis is another an extension of this externalist idea—a person’s mental life can outlast her. This is because whether we think at a time *can* depend on how we are related to things and events at that time, rather than on how we are intrinsically at that time. In what follows, I shall further spell out this view, provide my main argument for it, and detail some its advantages.

3. Caught in a web

Despite the thesis of this paper, I accept that often when someone thinks at a time, they do so in part because of some way they are intrinsically at that time (indeed, this is true of all typical cases). Let us say that a person *houses* a thought at a time just in case she has this thought at that time at least in part because of some way she is intrinsically at that time. Such housed thoughts may overflow a person’s spatio-temporal boundaries in *some* ways—the contents of such thoughts may be partly determined by factors external to her (as per externalism about mental content), or the fact that she has these thoughts at all may rest upon the workings of her extended mind and how it currently relates to her intrinsic properties (as per the extended mind hypothesis), or the types of attitudes in which her thoughts consist (i.e. whether they be beliefs, intentions, etc.) may rest in part upon her evolutionary history, etc.—but someone cannot have a *housed thought* at a time at which the person does not exist.

Note this, however: thoughts do not exist in a vacuum. That is, at least relatively mild holism about the mental obtains. According to this view, our thoughts (understood broadly, to

encompass beliefs, desires, emotions, sensations, intentions, other forms of mental activity, powers, etc.) form mental webs. Each thought is more or less strongly interrelated to other thoughts within such a web (though proponents of this mild holism need not insist that a thought gains its entire identity from its interrelations with the other thoughts in the web). On this view, it is fruitful to think of the mind not simply as a collection of individual thoughts, but as a *pattern* (or *patterns*) of thoughts.

On some stronger versions of such a holistic view of the mental, the idea that we may identify individual thoughts is itself a fiction. As David Lewis points out with respect to our representational mental states:

A serious issue, and one on which I take folk psychology to be agnostic, concerns the relation between the whole and the parts of a representation. Suppose I have a piece of paper according to which, *inter alia*, Collingwood is east of Fitzroy. Can I tear the paper up so that I get one snippet that has exactly the content that Collingwood is east of Fitzroy, nothing more and nothing less? If the paper is covered with writing, maybe I can; for maybe 'Collingwood is east of Fitzroy' is one of the sentences written there. But if the paper is a map, any snippet according to which Collingwood is east of Fitzroy will be a snippet according to which more is true besides...No snippet of a map is big enough that, determinately, something is true according to it, and also small enough that, determinately, nothing is true according to any smaller part of it. If mental representation is map-like...then 'beliefs' is a bogus plural. You have beliefs the way you have the blues, or the mumps, or the shivers. (Lewis 1999, 310-311)

In general, if our representational mental states are map-like, then there is a sense in which the whole is prior to, or at least on level pegging with, its parts. Indeed, Lewis sums up his position

at the end of this paper by saying ‘The contentful unit is the entire system of beliefs and desires.’ (Lewis 1999, 324), thus committing to a rather strong holism.

We need not take our holism so far. The pertinent point for our purposes is not that the mental whole is prior to its parts, but that individual thoughts (again, broadly understood) are related *systematically*, and that this being so is part of the essential character of minds like ours. One way to bring out this systematicity is to consider the ways in which changes to some aspect of the mind can have a significant ripple effect throughout the rest of the mind. Daniel Dennett considers the possibility of neurosurgeons implanting a lone belief in an agent, Tom (the belief being that Tom has an older brother in Cleveland). Dennett describes one possible outcome as follows:

Tom is at a party and in response to the question, 'Are you an only child?' he replies, 'I have an older brother in Cleveland.' When he is asked, 'What is his name?' Tom is baffled. Perhaps he says something like this: 'Wait a minute. Why do I think I have a brother? No name or face or experiences come to mind. Isn't that strange: for a moment I had this feeling of conviction that I had an older brother in Cleveland, but now that I think back on my childhood, I remember perfectly well I was an only child.' If Tom has come out of his brainwashing still predominantly rational, his induced belief can last only a moment once it is uncovered. (Dennett, 1981, 251)

A lone belief that is *not* systematically related to (in this case, largely by means of coherence *with*) other mental states cannot be sustained. Dennett describes two other possible outcomes of the neurosurgeons' experiment. The first is that Tom is given, alongside the belief that he has an older brother in Cleveland, a whole host of other beliefs that cohere with this belief, and many of Tom's original beliefs are erased. Indeed, we may extend this idea to include other kinds of

mental states that might relate to such a belief (desires, emotions, mental imaginings, etc). In such a case, the erroneous belief is sustained precisely by being part of a web of other mental states that cohere with each other. The last outcome Dennett imagines is of a deeply irrational and incoherent Tom maintaining his belief that he has an older brother despite also accepting that he is an only child, that he has no memory of such a person, that he has no desires relating to this person etc. Such a case is so bizarre that it is difficult to recognise as a genuine possibility. Even if it is, its unusual nature highlights the norm—that thoughts generally systematically relate to each other.

It is also worth noting that, given I am using the term “thought” to encompass any kind of mental activity or power, many thoughts essentially involve organised patterns of mental states. Mental systems, capacities, faculties, etc., which are concerned with the production, organization, change, or cessation of mental states (such include our faculties for memory, language, reason, and emotion) govern multiple mental subsystems and, ultimately, individual beliefs, memories, episodes of emotion, etc. Other faculties, like access consciousness (see Block 1990), have access to multiple such systems and the mental states to which they pertain. In essence, the mind is not simply a ragtag collection of singular mental events, but rather an organised pattern of interwoven mental states governed and utilised by multiple interrelated mental systems.

It is because of such facts that we may sensibly talk of thoughts making up mental webs—individual mental states are related by way of coherence, governance by the same mental systems, and other such relations. Enough such mental states, appropriately and strongly enough related, make up a mental web. Of course, it is difficult to specify exactly what relations must hold between thoughts for them to compose such a web (and how to determine the relevant

strength or degree of such relations), though I trust we all have a relatively good sense of what counts and what does not. As a referee points out, a certain type of rational connectedness is insufficient—consider a (schematic) case in which S believes *If P then Q*, T believes *P*, and U believes *Q*. Though such beliefs are rationally connected in the sense that the first two entail the third, there is no useful sense in which they make up a web (or even part of one). This is no doubt because, in such a case, there need be no underlying and appropriate causal or dispositional connection between any of these beliefs. Contrast such a case with one in which someone explicitly infers *Q* from *P* and *If P then Q*—in such a case, the rational connection between these beliefs is further reflected in the psycho-causal connections between them. Such a relation between beliefs clearly *is* of a type to render these beliefs part of the same mental web.

The holism I espouse here, then, is mild indeed—it simply highlights the fact that thoughts occur in an organised mental system of such thoughts, and thus bear various causal, dispositional, and psychological relations to other thoughts in such a system (I shall hereon often refer to such relations as *psychological* connections or relations). Thoughts which are interrelated by strong enough psychological connections make up *mental webs*.

One consequence of this view is that a person, in having a thought, is related to an entire mental web of thoughts that are intimately related to this thought. A person, in having a thought, has many thoughts. We may, then, take the primary mental relation between a thinker and her thoughts not as one between the thinker and each thought, but as one between a thinker and a mental web of thought. She has one thought by being appropriately related to a mental web which includes this thought.

Furthermore, we may think of mental webs *diachronically*. Some thoughts *inferred* from earlier thoughts are part of the same temporally extended mental web; so too are thoughts that

are appropriate *continuations* of earlier thoughts (for instance, if an unbroken sensation of pain occurs throughout a certain time, the later parts of this sensation are continuants of the earlier parts); so too are occurrent thoughts that are *manifestations* of earlier dispositional thoughts, etc. Thus, if a person has some thoughts at time t , and some later thoughts (which occur at t^*) are appropriately psychologically related to these former thoughts (such that they are part of the same mental web), then the person has these later thoughts too (*at t^**).

Given that one thinks the thoughts that occur in any mental web to which one is appropriately related, my thesis is that it is possible for a person's mental web to overflow her temporal boundaries. While I may have a certain temporal extent, some of the thoughts within one of my mental webs may occur beyond this extent. In other words, my thoughts may occur after (or perhaps before, though I shall concentrate on the former case) I exist.

This is possible if the type of psychological relations that tie thoughts together in the appropriate manner to form a mental web can *span objects*. And such relations *can* span objects. Indeed, this possibility is illustrated by many of the typical thought experiments around which the topic of personal identity revolves. Take, for instance, the idea that we can upload our mental lives to computers. Assuming this is possible (and that I undergo such a process), the thoughts that end up in the computer are appropriately connected to those earlier thoughts that were housed in *me* (they are continuations or manifestations of them, or inferred from them, etc.). These computer-housed thoughts are thus part of the same mental web as (some of) the thoughts that I once housed. Given this, I think the thoughts that are housed in the computer. And this remains so even if, after the upload, I am annihilated, leaving only the computer to house such thoughts. In such a case, I think the thoughts housed in the computer because these thoughts are appropriately psychologically connected to the past thoughts I housed when I *did* exist.

Perhaps one rejects the idea that our mental lives can be uploaded to computers—computers, no matter how sophisticated, just aren't the kinds of things that can house thoughts. Very well—consider instead your mentality being transferred to another organism (thanks to a sophisticated device that can read and extract brain patterns), or even your entire brain being thus transplanted. In such (appropriately described) cases, your thoughts (or many of them) survive the journey. If in this process you are destroyed, then *your thoughts outlive you*. You think these thoughts because they are part of your mental web (due to their close psychological connections to your previous thoughts).

My general case, then, for the possibility of someone's thinking without existing (at some particular time) is this. First, necessarily, if a thought, T, that occurs at a certain time, t, bears the appropriate psychological connections to thoughts housed by someone, S, then S has T at t (this being supported by the mild holism set out above). Second, it is possible for T to bear such appropriate psychological connections to other thoughts housed by S yet not itself be housed by S at t (this being supported by possible cases in which such psychological connections span between S and another being). And finally, if these first two claims are true, it is possible that S has T at t, despite S's not existing at t (this being supported by extensions of the previous cases in which S ceases to exist before t, but there remain cross-temporal psychological connections between S's past thoughts and the thoughts housed in the other being at t, including T).

A further consideration in support of the idea that we continue thinking when our mental webs outlast us worth highlighting is that the view makes good sense of a person's *fitting first personal concerns* in such cases. If a person's thoughts are appropriately continued in other beings (and these thoughts have the requisite psychological relations to thoughts she once housed), it is fitting for her, while she exists, to anticipate, fear, look forward to, welcome, etc.,

these thoughts to come. Thus, for example, if her thoughts are to be transferred to another body (through some sci-fi device, or a brain transplant, or what-have-you), and she knows that this new body is then to be tortured, it is fitting for her to fear the pain to come. But this would not be fitting unless the coming pain will be *her* pain.

4. Thought-transfers and person-transfers

It likely will be thought by some that I have taken the wrong lesson from those thought experiments which involve transfers of thoughts from one thing to another. I claim that such scenarios illustrate the possibility of someone's thinking a thought at a time without housing that thought at that time (and, thus, potentially without even existing at that time). More typically, such thought experiments (which date at least as far back as John Locke's famous prince and the pauper example (Locke 1975)) are used to support the idea that *people themselves* may be transported from one substance to another (again, Locke arguably being a proponent of this view). Why take the drastic measure of denying that thinking entails existing, when one can instead simply say that we are the kinds of beings that may swap bodies?

Though I am not very sympathetic with such views (because (a) I am rather convinced by the arguments for the thesis that we are organisms or bodies (see, e.g., Ayers 1991, Bailey 2016, Blatti 2012, Olson 2007) and (b) the view that we can think without existing can, in my view, better resolve certain puzzles of personal identity than the view that persons can transfer from one body to another (see sections 5, 6, and 7 for my treatment of such puzzles)), I wish here to present a conciliatory response. That is, I wish to show that my argument for the possibility of non-existent thinkers is perfectly compatible with the idea that, in standard thought-transfer cases, a person is transported from one body to another, and thus my argument may be accepted by those who hold this view. I have two arguments for this position. I present the first below.

Let us grant for the sake of argument, then, that a typical brain transplant, a complete mental transfer from one brain and body to another, or a total persona upload transfers a person from a body to another object (such as another body, or a computer). If this is right, then, in such cases, this person's mental web does not outlast the person, but rather *drags the person along with it*. Still, even if this is what happens in such cases, these scenarios are still cases in which someone thinks thoughts without housing these thoughts (and thus may think without existing). This is because, if one accepts the "person-transfer" account of these cases, one should also accept that there is more than one thinker present in such scenarios, and at least one such thinker is *not* dragged along with the mental web, but is instead *left behind*. At the very least, there is a strong case to be made for this claim.

A person's body thinks (perhaps a person's brain also thinks; perhaps various temporal or spatial parts of a person think too—if so, this is all the better for my point). After all, the body has the right kind of neural activity to think, it behaves like it thinks (it uses language, etc.), and there is a straightforward sense in which it houses mental states. Indeed, there seems little evidence that I think that is not also evidence that my *body* thinks. And, if it looks like a duck, swims like a duck, and quacks like a duck, it's a duck.

Furthermore, if a person can be transferred from her body to another, then she is not her body. Thus, on such a supposition, there are (at least) two thinkers that spatially coincide, at least for a time—a transferrable person and that person's body (Baker 2007 deems bodies that coincide with the people they constitute *derivative people*). And while the former may move from one thing to another in a typical thought-transfer scenario, the latter does *not* do so (a body is not uploaded to a computer even if a person inhabiting that body is, for example).

These points being granted, my original argument straightforwardly applies to people's bodies. The thinking body in a thought-transfer case houses certain thoughts (at a time just before the transfer). These thoughts form a mental web together with thoughts had at times *after* the transfer. These later thoughts are not themselves housed in the body, but in a different body (or a computer, etc.). Thus, if the (original) body is obliterated during the thought-transfer, the body's mental web outlives it. As I have already argued, in such an eventuality, the body thinks thoughts after it has ceased to exist. This remains so even if another thinker transitions from inhabiting one body to inhabiting another. It is thus possible for someone to think at a time without existing at that time, even if we accept the possibility that *thought*-transfers are also *person*-transfers.

Let's move on to my second argument that those who embrace that full thought-transfers are also person-transfers may still accept my case for the possibility of non-existent thinkers. This second argument relies on the fact that, even assuming that persons *can* be transferred when thoughts are, thought-transfers are person-transfers only when a *significant amount* of thoughts are transferred from one object to another. Tales of body-swaps, teletransportation, computer uploads, and the rest are most plausibly conceived as person-transfers only when, according to these tales, the person's *entire persona* (or something near enough) is transferred from one thing to another. Marcela Herdova, who is an advocate of the idea that persons swap bodies in typical thought-swap cases, is particularly explicit about this:

[T]o elicit the "body swap" judgment, we should describe the case as a total information swap (or a near total information swap) where not simply A's memories and traits will be transferred to B's body, but also his specialized knowledge of British history, his ability to speak French, his holiday plans, his sexual orientation, as well as his mental systems

and his first-person perspective. Given a (near) total information swap, many people are likely to judge that A and B have swapped bodies. ...Only then do we arrive at a convincing case of a body swap. (Herdova 2016, 328)

It is, of course, difficult for proponents of the possibility of person-transfers to say how much of a person's mentality must be transferred from one thing to another before we consider the person herself to have also been transferred. It is even difficult to say exactly how we might *quantify* a person's mentality. Regardless of how we might attempt such things, the point remains that there is some significant quantity of a person's mentality (however this quantity is conceived) the transfer of which does not suffice for the person's being transferred.

There are, what's more, possible cases in which a transfer of thoughts (i.e., the spanning of psychological connections between objects) is highly limited—perhaps only certain continuous conscious episodes are transferred, for instance. Such cases would fail to meet the threshold for a person-transfer.

If all this is right, my argument for the possibility of non-existent thinkers can be run using these highly limited transfers. Consider, for instance, a conscious mental episode being transferred from one body to another. This conscious episode alone falls below any threshold for being a person-transfer. Still, in such a case, the person who houses the beginning of this conscious episode undergoes all of it because there is the appropriate connection between the start of the episode and the rest of it—the start and the end of the episode are both parts of this person's mental web. Seeing as the person does not house the end of the episode, she undergoes it in virtue not of how she is intrinsically at the time of its occurrence, but how her earlier mentality relates to it. Given this, should she cease to exist in between the time she houses the start of the conscious episode and the time the end of the episode occurs, she undergoes the end

of the episode at a time at which she does not exist. Her mental web overflows her temporal boundaries, if only by a smidge.

I conclude, then, that the case for possible non-existent thinkers is perfectly compatible with the common view that typical thought-transfer scenarios are also *person*-transfer scenarios. Of course, I do not consider the above support for this to be conclusive (what in philosophy is?), but I take it to show, at the very least, that my basic argument for the possibility of thinking without existing is in no great tension with a common view that, at first blush, may seem in deep conflict with it. In what follows, I shall address how my thesis provides an attractive solution to certain puzzles in personal identity.

5. Fission cases

Fission cases involve a psychological (or sometimes physical) split in a person. One classic fission case was proposed by Wiggins (1967, p50) and developed by Parfit (1971, 1984). Here is my (slightly altered) version of this case:

Brain transplants have been perfected. Indeed, it is now possible for surgeons to transplant only half of one's brain into a new body and retain nearly everything that is distinctive about one's mentality (one's personality, beliefs, intentions, etc.). That is, the person whose half-brain is removed is *psychologically connected* with the person who receives this half-brain. Unfortunately, I am captured and placed in room A, in which I am told, at time t1, what is about to happen. After t1, my brain is bisected and removed, and one half is placed in a new body in room B, while the other is placed in a new body in room C (my body is then destroyed). Both operations are successful. The person who is in room A at t1 (i.e., me) is psychologically connected with *both* of the people who

wake up after t_1 in rooms B and C. Both of these people (those in rooms B and C) are painfully tortured at t_2 .

The above scenario, although not (currently) *technically* possible, is nonetheless *metaphysically* possible.

It has long been recognised that the case above is puzzling. On the one hand, there seems to be an excellent case that I no longer exist at t_2 —the people who are located in rooms B and C at t_2 are distinct people from each other, thus I am not identical with both of them, but nor am I identical with only one of them, as any (identity-relevant) relations that I bear to one of them I also bear to the other. If I am neither of the people in rooms B or C at t_2 , I do not exist at t_2 . On the other hand, it seems that I have good prudential reasons to anticipate and fear (at t_1) being in agony at t_2 , which seems like excellent evidence that I *do* exist at t_2 —If I rationally fear at t_1 being in pain at t_2 , then I am in pain at t_2 . And if I am in pain at t_2 , then I exist at t_2 .

I endorse the bulk of the above reasoning except the last step: we illicitly infer from the fact that I will be in pain at a certain time that I exist at this time. It is this inference I deny, and, indeed, this inference that gets us into trouble. We can accommodate all the seemingly conflicting claims we wish to make about the above fission case (and others), simply by giving up the idea that thinking at a time requires existing at that time. That is, we can endorse both the quite compelling case for the idea that I no longer exist at t_2 , and the just as compelling case that I will be in pain at t_2 . What's more, we can provide a rather nice *explanation* of how it can be that I feel pain at a time at which I do not exist: though I no longer exist at t_2 (having been destroyed), I count as being in pain at t_2 because my earlier mental states bear the appropriate psychological connections to the pain that occurs at t_2 .

It is worth comparing my treatment of fission cases with Derek Parfit's (as set out in his classic 1971 paper (see also his 1984)). Parfit takes fission cases to show that it is possible to have appropriate first-personal concern for people who are not me, but are rather my "psychological descendants" (i.e., those future people who, whether they are identical with me or not, are psychologically connected with me). For Parfit, it is appropriate to fear the future agony not because *I* will undergo it, but because my *psychological descendants* will undergo it. My having psychological descendants is *what matters* as far as my survival concerned—my continued existence, on the other hand, does *not* matter. Survival, as the saying goes, comes apart from identity.

Parfit and I agree on much. He too thinks that I no longer exist at t_2 . He believes, as do I, that it is appropriate for my first-personal concern to extend to these times, even though I do not exist at these times. In particular, he would be happy to concur that my fearing at t_1 the agony at t_2 would be perfectly appropriate. We also agree that I am psychologically connected to two different future people. However, Parfit thinks that, therefore, my (fitting) concern about the coming torture is concern for the pain of those people. Those people are not me and thus, on Parfit's view, one can rationally have first-personal concern for people other than oneself.

This line of reasoning, though attractive, is too quick to give up on the idea that my fitting first-personal future-directed attitudes are about *me*. Indeed, the claim that it is appropriate for me to fear being in agony and the claim that it is appropriate for me to fear *my* being in agony strike me as (necessarily) equivalent. If I were to be truly convinced that I would feel no pain at all at t_2 , I would not fear any such future pain. I of course may fear *for* those whom I understand will be in agony, but fearing and fearing-*for* are different attitudes, governed by different norms.

But if I am right about this, where does Parfit's reasoning go wrong? Though Parfit's reasoning seems initially compelling, he has overlooked the following possibility—my psychological connection to the people in agony in rooms B and C at t2 not only grounds the appropriateness of my fearing at t1 the future agony, it also grounds the fact that *I will be in agony at t2*. That is, it is not only the people in rooms B and C that will be in agony at t2—I too will be in agony. I will be in agony *because* I am psychologically linked to that agony in an appropriate way. Though I may not *be* either of the people in rooms B and C after t1, I share their pain. Given this possibility, Parfit has failed to sever the link between my (fitting) first-personal fear and *me*. In essence, we may accept Parfit's reasoning that I no longer exist after t1, *and* that I have appropriate first-personal concerns directed at times after t1, *without* accepting that these first-personal concerns are about *other people*. We may instead maintain that these concerns are directed at *me*, and in particular, at what I will think and feel after t1.

We may reiterate this point by observing that there is a crucial ambiguity in the claim that identity is what matters in survival (and thus in the corresponding claim that identity is *not* what matters). On the one hand, the claim might be that what concerns us (and what should concern us) when considering our survival is our *continued existence over time* (plus a certain quality of mental life). When I look forward to the future, what is important for me is that *I* exist in the future, and enjoy a certain quality of life. On the other hand, the claim might be that what concerns us (and what should concern us) when considering our survival is that *we go on thinking* in the future (and having a certain quality of life). When I look forward, what is important is that I *think* in the future (and have a certain quality of life). It is easy to conflate these two claims, but they are distinct; denying the first is not equivalent to denying the second.

To my mind, Parfit has persuasively argued that my continuing to exist is not what matters for my survival, but he does not succeed in showing that my continuing *to think* is not what matters.

A final point is worth noting—my picture solves a residual puzzle for another proposed solution to the fission problem. Some views (such as that of Lewis 1976, who spells out this view using his perdurantist metaphysics) suggest that, in fission cases, two people coincide pre-split, and no longer coincide post-split. Even if this is right, each person pre-split is strongly psychologically connected to *both* of the people post-split. Someone with Lewis's view, then, is left with the question of what kind of *pre-split* first personal concern each such person should have toward the mental well-being of the post-split person to which she is *not* identical. On the one hand, she is *not* this person, which suggests she should not have distinctive first-person concerns towards this person's thinking. On the other, she *is* strongly psychologically connected to this person, which suggests she *should* have such concern. This is, in effect, a close cousin of the puzzle that fission cases initially present. My view solves it in a similar way—even if one of the pre-split persons is (exactly) one of the post-split persons, the post-split person she is *not* is strongly enough psychologically connected to her that they share mental states. It is thus fitting for each pre-split person to have first-personal concern for the mental well-being of *both* post-split persons—fission cases are ones in which people's mental lives overspill their spatiotemporal boundaries.

6. Williams's redescription puzzle

Usual descriptions of thought-swap cases elicit (or are at least designed to elicit) the view that people can swap bodies. Thus consider a case in which two people have their mental information extracted from their bodies, and the mental information from one body is put in the other (and vice versa). We may vividly describe such a case as a person's being in one body then waking up

in another, confused as to their predicament. Bernard Williams (in his classic 1970 paper) further imagines that two people are asked, pre-swap, which of the post-swap bodies they wish to receive money, and which tortured. It seems the self-interested choice is to ask for the body one currently occupies to be tortured and for the other body to receive money. This, of course, suggests that such cases truly are body-swap cases.

Williams points out, however, that such cases can be redescribed to induce the judgment that persons in such cases do not swap bodies. Rather, we may see these scenarios as ones in which such people undergo radical mental changes to *become mentally like the other person was*. Williams produces this effect by taking us stepwise through a series of smaller changes someone might undergo:

- (i) A is subjected to an operation which produces total amnesia;
- (ii) amnesia is produced in A, and other interference leads to certain changes in his character;
- (iii) changes in his character are produced, and at the same time certain illusory “memory” beliefs are induced in him; these are of a quite fictitious kind and do not fit the life of any actual person;
- (iv) the same as (iii), except that both the character traits and the “memory” impressions are designed to be appropriate to another actual person, B;
- (v) the same as (iv), except that the result is produced by putting the information into A from the brain of B, by a method which leaves B the same as he was before;

(vi) the same happens to A as in (v), but B is not left the same, since a similar operation is conducted in the reverse direction. (1970, 172)

Suppose, as Williams does, that A's body will be tortured after being mentally altered, and A is informed of this before any of these alterations. Williams asks us whether, for each of (i)-(vi), it would be appropriate for A to fear the coming torture, given A knows what mental alterations he will first undergo. The first step, (i), is surely simple: just because A will first lose his memories will do nothing to stop the agony of the torture to come, and thus it would still be fitting for A to fear the torture. But if that is right, the difference between (i) and (ii) doesn't seem to make such fear any less fitting. Proceeding in this fashion, Williams argues that, even by step (vi), it would still be fitting for A to fear the coming torture. But notice this: step (vi) is simply a redescription of our initial supposed body-swap case. Framed in one way, it is intuitive that the scenario is one in which people swap bodies. Framed in another, it is intuitive that the scenario is one in which people do *not* swap bodies.

How do we account for this? Herdova 2016 provides a taxonomy of possible answers to this question: the "Classic View" insists that the body-swap judgment about such cases is correct; the "Dismissive View" tells us to doubt either the possibility of such cases or our judgments about such cases; the "Bold View" tells us that animalist judgments about the cases are correct (i.e., judgments that the people stay in their bodies); and the "Equivocal View" tells us that our judgments about such cases are equivocal, which may lead to the Dismissive View, or alternatively to the view that facts about personal identity are indeterminate.

All these views attribute some major error to our thinking about these cases—we are either wrongheaded in our judgments when the cases are framed in one way, or when they are

framed in the other way, or when framed in *both* ways. My picture suggests an alternative according to which our judgments under either framing are tracking significant truths.

First, then, consider the usual (non-Williams-style) framing. We judge that the person whose pre-swap mental information ends up in the tortured body (call her “A”) has reason (pre-swap) to fear the torture—this reason being that it is she who will be in terrible pain because of this torture. My picture accommodates this judgment by agreeing—A *will* be in terrible pain because of this torture, and thus it is fitting for her to fear the torture. Furthermore, the explanation my view provides for this claim is essentially the same as a typical psychological continuity theorist provides—A’s mentality is significantly psychologically connected to the mentality of the tortured post-swap body, which suffices for A’s thinking and feeling what the tortured person in the post-swap body thinks and feels. *Unlike* psychological continuity views, my view stops short of claiming that A herself changes from being located in her pre-swap body to inhabiting the tortured post-swap body. Rather, she continues to think and feel what this post-swap body feels, even without inhabiting it, because of how her pre-swap mentality is related to the mentality of the post-swap body. Nonetheless, the pain she will experience will truly be *her* pain, and thus something for her to fear. The intuitive body-swap judgments we make capture this highly significant point, while bringing with them a natural but mistaken internalist assumption that A’s feeling the pain must mean she (eventually) inhabits the tortured body. While this is indeed a mistake on my picture, it is both understandable and, practically-speaking, relatively minor; what is of much greater importance is that A will feel the pain and is correct to fear it—*this* the body-swap judgment gets right.

Turn now to the Williams-style framing of the case. Once in the mindset Williams intends to bring out, we judge that an agent, let’s call her “B”, stays in her body—her

anticipation and fear of the pain to come can “see through” the massive psychological changes she will undergo. This is so even though such changes make her psychologically just like A, and A in turn will be made just like B. My picture accommodates this judgment by agreeing—B *will* be in terrible pain because of this torture, and thus it is fitting for her to fear the torture. Furthermore, the explanation my view provides for this claim is essentially the same as a typical bodily continuity theorist provides—B remains in her body (perhaps *is* her body, though we need not commit to this much) and thus will undergo any torture that is meted out to her body. While B’s pre-swap mentality will not be psychologically connected to her post-swap mentality, she will still feel the pain of the torture because *she will inhabit the body that will be tortured*. The Williams-style judgments we make capture this highly significant point, though they may bring with them the natural but mistaken internalist assumption that B should not have first-personal concern for the mentality that is transferred from her body to A’s body.

The upshot is that both framings get important things right. As per the usual framing of swap cases, someone whose persona is transferred from her body to another (and receives the persona from that body) does go on thinking in this new body, due to the strong intertemporal psychological relations between the mental states of each of these bodies. It is thus also fitting for her to anticipate this thinking in a distinctly first-personal manner (“What will I think of this new body?”, “I hope it won’t hurt too much”, “I wonder whether this other body will feel different than mine”, etc.). All of this is predicted by both the standard psychological views and the picture I am advocating in this paper. Standard animalist and body views, however, have difficulty accommodating such a verdict. On the other hand, as per the Williams-style framing of swap cases, someone whose body receives a new persona from another body (while her persona is placed in this other body) will go on to have the thoughts and experience that occur in her

original body, including the pain of being tortured. This is predicted by animalist and body views, and the picture I am advocating, while psychological continuity views have trouble accommodating such a verdict.

How powerful the above considerations are may in the end depend on how much store one puts in our judgments about thought-swap cases. I have argued that my theory accommodates the most important such judgments even given the radically different ways in which the cases can be framed. I take this to be a considerable benefit of my view. If one rejects some of these judgments as misleading or off-base, one may instead see it as a cost. This said, I think people's tendencies to dismiss such judgments come from the understandable but mistaken fear that not all of these important judgments *can* be suitably accommodated, and thus they feel forced to reject them. Part of my point is that this is not so.

7. Thinking by degrees

My picture is consonant with, and can be further informed by, another of Parfit's arguments that survival comes apart from identity—that survival comes in degrees while identity does not. The latter claim is relatively straightforward—on standard views there is no such notion of being *a bit identical* or *identical to some degree* to something (there is, of course, the notion of *partial identity*, which some say is the relation between two mereologically overlapping objects, but such a notion is seldom taken to be literally the identity relation). Parfit argues for the former claim with reference to further puzzle cases—a fusion case in which two people merge mentalities, and a case in which certain creatures regularly undergo fission (1971, p18-21). Do the people who undergo fusion survive? Well, yes, to some degree—they each bear certain strong psychological connections to the newly created fused entity, but the fused entity is also radically different from each of them in other ways. The creatures that regularly divide

eventually have descendants that are mentally quite distinct from them. The degree to which the original creatures survive, then, decreases with each division.

On my picture, talk of survival coming in degrees translates to talk of the degree to which a person has certain thoughts. Consider again Parfit's dividing creatures. One original such creature can think for some while without existing as long as her mental web stretches far enough out ahead. But how far a mental web stretches is a matter of how strong the psychological connections are between the thoughts of the original being and the thoughts of its descendants (because a mental web is made up of thoughts which are sufficiently psychologically related to each other). Such strength of connection comes in degrees. In a perfectly comprehensible sense, then, we may say that *having a thought* can come in degrees. Roughly speaking, one has a thought to a degree lesser than the highest possible degree just in case this thought is less strongly connected than is possible to the thoughts one houses. (In fact, we may want to sophisticate this account to privilege some thoughts we house over other thoughts we house, but I shall ignore this complication here.) If the strength of the psychological connection between a thought and the thoughts housed by a person dips below a certain threshold (a threshold I shall not specify here because I do not know how to go about doing so), the person does not have this thought.

Indeed, the same basic idea applies to the fission case discussed in section 5—as the people to whom I have given half my brain each go on living, and changing, the thoughts they have will become more and more distantly related to the thoughts I once housed. Eventually, given enough psychological distance, I no longer think the thoughts they house at all.

We may put the point in terms of mental webs. A mental web is made up of a variety of interconnected thoughts, spread across time (and space). We should not suppose, however, that

“being in the same mental web as” is transitive. A mental web is a web of mental states that *are all sufficiently psychologically related to each other*. On such an understanding, it is perfectly possible for two states, S and T, to be in the same mental web as each other, while only the latter is in another mental web with other thoughts. To see this, simply concoct a scenario in which T is on the periphery of the second mental web, but is still just about sufficiently related to the other mental states in it, and S is less intimately related to the other mental states in this web (memories may provide an interesting example of this phenomenon—consider a case in which an agent at one time has access to, and often makes mental use of, two separate memories, but, at time marches on, finds these memories harder to access, and makes less and less use of them; there may come at time at which one such memory is no longer part of any mental web the agent houses, while the other memory remains just accessible enough to count as part of some such web). Indeed, a person throughout their lifetime may house multiple overlapping mental webs (think of them like overlapping Venn diagrams). This person thinks a thought if it appears in any such mental web. Once she stops existing, however, she can no longer house any mental webs, and thus the last thoughts she has must appear in mental webs she once housed. This means such thoughts need to be rather robustly related to the thoughts she had when she *did* exist. As these mental webs change and develop, they are replaced with new such webs which the person never housed and thus the non-existent person stops thinking. Once one thinks without existing, one is destined not to burn out, but to fade away.

8. Concluding thoughts

My main aim in this paper has been to get a view that is otherwise dismissed (or, more accurately, ignored) on the philosophical map. Once one embraces an externalist, mildly holistic view of the mental, the position that a person’s thoughts might continue without the person is not

the incoherent claim one might initially think it is, but rather a natural outcome of well-supported and common views in the philosophy of mind. Furthermore, this position solves various difficult puzzles in the personal identity literature.

This is not to say, of course, that no questions remain. One might wonder, for instance, what else one can do without existing. Is one condemned simply to think without perceiving the world, or acting on it? Or might one not only think, but also see and hear, and interact with the world? In my view, a non-existent person may not only think, but also perceive and act. These latter two phenomena are, after all, mental (at least broadly speaking), and thus the considerations I present above for the idea that one might think without existing also apply to acting and perceiving.

Consider a case in which my mentality is put into a body while I am destroyed. On my view, I no longer exist, but I do think (thanks to my connection to this new body). Let's say this body is tortured but manages to escape by running away. Do *I* also run away? I think so. I intend to run away (this intention is strongly psychologically connected to the mental states I had pre-death), and this intention non-deviantly causes the action of running away. Thus, I run away.

One might object that running away consists in the movement of arms and legs, but I *have* no arms or legs. In general, acting involves *moving*. Surely, I cannot move at a time without existing at that time! This objection involves a subtle equivocation concerning the term "move". As Hornsby (1980) notes, "move" can be understood transitively or intransitively—on the former understanding it takes an object, and on the latter it does not. Armed with this distinction, we can say that, in the case under consideration, I move *something* (the other person's body), even though I do not myself *move*. I move something because my mental states are appropriately

causally related to the movement of the other person's body. My *moving something* in this manner suffices for my *acting*.

Similar considerations suggest that we may perceive the external world without existing. Continuing the above example, after I run away, I may direct the body I am controlling to turn on the television. Light travels into the body's eyes from the TV, sending signals to its brain, which then causes it to have the experience of watching television. Because I am strongly psychologically connected to these experiences, they become part of *my* mental web. And because of this, these experiences are mine.

The last point I shall mention is that my picture undercuts several arguments against animalist and body views of personal identity. Such arguments proceed, roughly, by highlighting cases in which someone seems to continue thinking even though her body no longer exists (or in which the body no longer houses thoughts, or no longer houses thoughts that are psychologically connected to the person's current thoughts). From this it is concluded that the person cannot be identical to her body because, after all, she continues to think while her body no longer even exists (or exists, but does not house thoughts, etc.). Though such examples date back to the dawn of the personal identity literature (i.e., Locke 1975, though see also Parfit 1971, 1984, Shoemaker 2011, Baker 2007, Herdova 2016), it will suffice to mention only one case, due to Matt Duncan (see his 2015). Duncan asks us to imagine that, for a period of time, we contemplate the fact that $2+2=4$. The case continues thus:

While you are lying there thinking, aliens destroy most of your body. They do this sometime between t and t^* , and the only part of you that they don't destroy is your cerebrum. In fact, the aliens manage to sustain the normal functioning of your cerebrum.

So, while you lose most of your body sometime between t and t , you still think ‘ $2+2=4$.’
(Duncan 2015, 750-751)

It’s easy to see where Duncan wants to lead us—if we are still thinking after our bodies are destroyed, after there is no animal left, then we are not our bodies, nor are we (essentially) animals. Given my line in this paper, it should also be easy to see where, according to my picture, Duncan goes wrong. We cannot conclude from the fact that my body no longer exists, plus the fact that I am still thinking, that I am not my body. This is because it is possible for me to think without existing. If I am my body, then, we simply have a case in which my body is thinking without existing. And indeed, Duncan’s case is plausibly one such case—though I may be the subject of a stream of thought that lasts from one time to another (i.e., I think this thought throughout this period), there is a point at which I stop *housing* this thought (and, indeed, at which I stop existing). Though I may no longer house the stream of thought, however, I still have it because it is a direct continuation of the stream I *did* house. In general, given the possibility of something’s thinking without existing, this standard kind of criticism of animalism and body views fails. Note, however, that the kind of animalism that can resist this kind of case has taken on board many of insights of typical psychological continuity views. In a way, my picture is a *synthesis* of the insights of bodily and psychological continuity views.

My hopes are not so high as to have convinced you that it is indeed possible to think without existing (though that would be nice!). I do hope, however, to have persuaded you that the notion is by no means as absurd as it first appears. Indeed, the view is harmonious with popular and plausible externalist views of the mind, provides elegant solutions to various puzzles of personal identity, provides animalists with a strong line of defence against the most prominent attacks on it, and follows from rather modest commitments (such as to the mild holism of the

mental). While I may not have turned your incredulous stare into a nod of agreement, perhaps you may grant me a quizzical shrug.

Acknowledgements

Many thanks to an anonymous reviewer for very helpful comments, and to attendants of the 2017 Undergraduate Conference for Philosophy and Theory at the University of Florida, who heard ancestors of some of the ideas in this paper and were kind enough to humour them.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Ayers, M. (1991) *Locke: Epistemology and Ontology*. Routledge: New York.

Bailey, A. (2016) "You Are an Animal." *Res Philosophica*: 93: 205–21.

Baker, L. R. (2007) *The Metaphysics of Everyday Life: An Essay in Practical Realism*.

Cambridge University Press: Cambridge.

Blatti, S. (2012) "A new argument for animalism." *Analysis* 72:685-690.

Block, N. (1990) "Consciousness and Accessibility." *Behavioral and Brain Sciences* 13: 596-598.

Burge, T. (1979) "Individualism and the mental." *Midwest Studies in Philosophy* 4: 73-122.

Clark, A. and Chalmers, D. (1998) "The Extended Mind." *Analysis* 58: 7-19.

Davidson, D. (1987) "Knowing One's Own Mind." *Proceedings and Addresses of the American*

Philosophical Association, 60: 441-58.

Dennett, D. (1981) "Mechanism and Responsibility." in *Brainstorms: Philosophical Essays on*

Mind and Psychology. Brighton: Harvester Press. 233-255.

Descartes, R. (1993) *Meditations on First Philosophy: Third Edition* (trans. D.A.Cress). Hackett:

Indianapolis.

Duncan, M. (2015) "I Think, Therefore I Persist." *Australasian Journal of Philosophy*, 93: 740-

756.

Herdova, M. (2016) "Nothing to Fear: Swap Cases and Personal Identity." *Analytic Philosophy*

57:315-337.

Herdova, M. (2021) "The Importance of Being Ernie." *Thought* 10: 257-263.

Hinton, J.M. (1967) "Visual Experiences." *Mind*, 76: 217-227.

Hornsby, J. (1980) *Actions*. Routledge & Kegan Paul: London.

Kearns, S. and Magidor, O. (2012) "Semantic Sovereignty." *Philosophy and Phenomenological*

Research 85: 322-350.

Kripke, S. (1982) *Wittgenstein on Rules and Private Language*. Blackwell: Oxford.

Langton, R and Lewis, D. (1998) "Defining 'Intrinsic'." *Philosophy and Phenomenological*

Research 58:333-345.

Leibowitz, U. D. (*forthcoming*) "Descartes' foundation and Borges' ruins: how to

doubt the *Cogito*." *Inquiry*.

Lewis, D. (1976) "Survival and Identity." in *The Identities of Persons*, Amelie Oksenberg Rorty

(ed.), 17-40, University of California Press: Berkeley.

Lewis, D. (1999) "Reduction of Mind." in *Papers in Metaphysics and Epistemology*. Cambridge:

Cambridge University Press. 291-324.

Locke, J. (1975) *An Essay Concerning Human Understanding*, Peter H. Nidditch (ed.). Oxford

University Press: Oxford.

Mele, A (2003) *Motivation and Agency*. Oxford University Press: Oxford.

Millikan, R. (1990) "Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox." *The*

Philosophical Review: 99: 323-353.

Olson, E. (2007) *What Are We?: A Study in Personal Ontology*. Oxford University Press:

Oxford.

Parfit, D. (1970) "Personal Identity." *Philosophical Review* 80:3-27.

Parfit, D. (1984) *Reasons and Persons*. Oxford University Press: Oxford.

Putnam, H. (1975) "The Meaning of Meaning," *Philosophical Papers, Vol. II : Mind, Language,*

and Reality. Cambridge University Press: Cambridge.

Shoemaker, S. (2011) "On What We Are" in *The Oxford Handbook of the Self*, S. Gallagher

(ed.). Oxford University Press: Oxford.

Waller, R. (2014) "The Threat of Effective Intentions to Moral Responsibility in the Zygote Argument." *Philosophia* 42, 209-222.

Wiggins, D. (1967) *Identity and Spatio-Temporal Continuity*. Blackwell: Oxford.

Williams, B. (1970) "The Self and the Future." *Philosophical Review* 79: 61-180.

Williamson, T. (2000) *Knowledge and its Limits*. Oxford University Press: Oxford.

Williamson, T. (2015) *Modal Logic as Metaphysics*. Oxford University Press: Oxford.