1

Introduction

1.1 What Is This Book About?

Here is the short answer. This book is about the metaphysics of personal identity, and the metaphysics of personal *ontology* in particular, where personal ontology concerns the question "What are we?" Over the course of the book, I will argue that it is much harder to determine which account of personal ontology is correct than many philosophers suppose. In the final two chapters, I will explore whether/how my arguments in previous parts of the book should impact our views regarding the possibility of life after death.

That's the short introduction. Here is a more detailed introduction.

This book concerns the *metaphysics* of personal identity. Questions regarding the metaphysics of personal identity are distinct from questions regarding, say, the psychology or sociology of "personal identity" – that is, the manner in which we conceive of the story of our lives, or the question of how we relate to other individuals and groups. There are two main questions which generally concern philosophers when they inquire into the metaphysics of personal identity: (1) Under what conditions is someone at some time numerically identical[1] with something at some other time? (2) What are we? The first of these questions concerns the nature of personal identity over time, while the second question concerns what philosophers call "personal ontology."

The first question, regarding personal identity over time, is the question regarding personal identity which is more commonly discussed among philosophers. It will prove helpful to review some prominent answers to this question, to give a sense of what question is being asked, and to get a

[1] Numerical identity is the relation that everything bears to itself and nothing else. To say, e.g., that Peter Parker is numerically identical with Spider-Man is to say that Peter Parker *is* Spider-Man. Numerical identity is contrasted with qualitative identity, where something is qualitatively identical with something else if and only if they share all of the same properties. Throughout this book when I write of "identity," I have in mind *numerical* identity, unless I say otherwise.

sense of how this question (and its most prominent answers) differs from the question regarding personal ontology (and its most prominent answers). Some prominent answers to the first question (regarding personal identity over time) are the following:

- *The psychological continuity view*: Someone at some time is identical with something at some other time if and only if they are suitably psychologically related to one another, where by "psychologically related" is usually meant "shares psychological states (e.g., memories, beliefs, desires)" or "linked by a chain of overlapping psychological states."[2]
- *The physical continuity view*: Someone at some time is identical with something at some other time if and only if they are suitably physically related (e.g., they are suitably biologically related), where the physical relation in question does not have a psychological component.[3]
- *The mixed view*: Someone at some time is identical with something at some other time if and only if they are suitably related by some mixture of psychological and physical continuity.[4]
- *The soul continuity view*: Someone at some time is identical with something at some other time if and only if they have the same soul. Here, "soul" usually means an immaterial thinking substance.[5] But sometimes the word "soul" is meant to refer to the "form" of one's body or the matter making up one's body.[6]
- *Anticriterialism*: People persist over time, but there are no informative necessary and/or sufficient conditions for when someone at some time is identical with something at some other time.[7]

Some prominent answers to the second question regarding the metaphysics of personal identity (the question regarding personal ontology, "What are we?") are the following:

- *Animalism*: We are animals.[8]
- *The brain view*: We are brains,[9] or particular parts of brains (e.g., cerebral hemispheres).[10]

[2] Locke's *An Essay Concerning Human Understanding*, Ch. 27 (Locke 1997: 304–305); Lewis 1976; Parfit 1984: §78; Shoemaker 1984; Noonan 2003.
[3] Williams 1970; van Inwagen 1990; Olson 1997; DeGrazia 2005.
[4] Nozick 1981: Ch. 1.
[5] As in Swinburne 1986, 2013, 2019.
[6] As in Stump 1995.
[7] Swinburne 1984; Lowe 1996: Ch. 2; Merricks 1998; Langford 2017.
[8] Van Inwagen 1990; Olson 1997; Snowdon 2014; Bailey 2015; Bailey and van Elswyk 2021.
[9] Parfit 2012. Maybe Nagel 1986: Ch. 3, §3.
[10] Puccetti 1973.

- *Constitutionalism*: We are physical objects "constituted" by, but not identical with, our bodies.[11]
- *The soul view*: We are immaterial souls.[12] Some of those who think that we are immaterial souls think that *everything* is immaterial. But most of those who think that we are immaterial souls think that some things, such as our bodies, are material, while other things, such as our souls, are immaterial. Those who endorse this latter thesis are known as *substance dualists*. (Here, "material" is synonymous with "physical," and the two terms will be used interchangeably throughout this book. "Immaterial" and "nonphysical" will also be used interchangeably.)
- *The soul+body view*: We are composites of souls and bodies.[13] This view differs from the soul view by claiming that we are not souls, although we are (currently) composed of an immaterial soul and a material body.[14]
- *The bundle view*: We are "bundles" of mental states.[15]
- *The nonself view*: "We" aren't anything, because we don't exist.[16]

The two questions regarding the metaphysics of personal identity are related, and the answer which one gives to one question will have implications for the answer which one gives to the other question. For example, if you think that strictly speaking persons do not exist, then of course you will not think that there are any conditions under which a person at some time is identical with someone at some other time. Similarly, if you think that there aren't any such things as souls (and so that *we* are not souls), then you will not think that the conditions under which a person at some time is identical with someone at some other time have anything to do with whether they have the same soul.

This book primarily concerns the second question regarding the metaphysics of personal identity, the question of *personal ontology*,

[11] Shoemaker 1984: 112–114, 1999, 2008b; Johnston 1987; Baker 2000.
[12] This view has been endorsed by a number of prominent philosophers, including Plato (*Phaedo*), Descartes (*Meditations on First Philosophy*), and Leibniz (*Monadology*). Some of its recent defenders include Foster 1991; Plantinga 2006; Unger 2006: Ch. 7.
[13] Augustine (*The Trinity*, XV.ii.11); Aquinas (*Summa theologica*, I, q. 75, a. 4); Swinburne 1986, 2013, 2019.
[14] To say that some *x*s compose a *y* is to say that the *x*s are all parts of *y*, and *y* has no other parts not included in the *x*s.
[15] This view might be endorsed by Hume in *A Treatise of Human Nature*, Book I, Part 4, §6 (Hume 2000: 165). See also Quinton 1962.
[16] This view is endorsed by many in the Buddhist philosophical tradition, e.g., the Pāli Canon's Anattalakkhaṇa Sutta (Bodhi 2000: 902), although often with the qualification that persons or selves exist "conventionally." For details, see Chapter 6 of this book. The nonself view is also endorsed by: maybe Hume's *A Treatise of Human Nature*, Book I, Part 4, §6 (Hume 2000: 165); Unger 1979a, 1979b; Rosen and Dorr 2002: §6; Sider 2013: §7.

"What are we?" But what exactly are we asking when we ask that question? In response to the question "What are we?" you might say, "We are things which are no taller than 10 meters." While it's true that we are things which are no taller than 10 meters, this doesn't really answer the question "What are we?" as that question is understood in debates regarding personal ontology. Debates regarding personal ontology are asking about our metaphysical nature. It is difficult to spell out what exactly that means (just as it is difficult to spell out what we mean by words like "metaphysical" and "nature"). The best way to get a grasp on what question is being asked here is to see some of the representative answers to that question, as we have just done.

Sometimes the question of personal ontology is put in terms of what we are *essentially*. This doesn't seem to me to be a helpful way to construe the question. Suppose, for example, that we are immaterial souls. Saying that we are immaterial souls would certainly answer the question "What are we?" as that question is understood in debates regarding personal ontology. But it does not automatically follow that we are *essentially* souls, in the sense that it is metaphysically impossible for any of us to not be souls. Someone might very well claim, rightly or wrongly, that something which is an immaterial soul is, in some other possible world, or at some other time, something other than an immaterial soul (say, a physical object). So, answering the question "What are we?" does not automatically answer the question "What are we essentially?"

Who is the "we" in the question "What are we?"? The individuals I have in mind are those living human individuals reading this book, as well as all those living human individuals who won't read this book. To say that we are concerned with the question of "personal" ontology may be misleading, since it gives the impression that the question which interests us has something to do with the notion of personhood, and the "we" in the question "What are we?" concerns all and only persons. But that's not right. For one thing, the "we" in the question "What we are?" might include persons as well as nonpersons. For example, suppose that at some point in the future I will exist in a vegetative state. In that case I might not be a "person," in some particular way of understanding the term "person." Nevertheless, I still intend the question "What are we?" to concern myself when I am in a vegetative state. On the other hand, the question "What are we?" is not meant to concern itself with many nonhuman individuals who *are* persons, or would be persons if they existed – for example, sentient computers, gods, angels, demons, ghosts, extraterrestrials. And there is no reason to think that an answer to the question "What are we?" will say what those *other* persons are. There is no reason to assume that all persons will

be the same sorts of things. For example, some persons might be material organisms, while other persons are immaterial beings.

A caveat: *Most* of this book will concern the question "What are we?" where "we" includes all and only living human individuals. But in the final two chapters, I turn my attention to the possibility of life after death, and some of what I say there will concern the personal ontology of beings who are not living human individuals. Some of what I say will concern *formerly* living human individuals – that is, human individuals who are dead. And the final chapter contains an extended discussion of "mind uploading," and so concerns the personal ontology of "uploaded" persons, and simulated persons more generally.[17]

I hope it's clear now what question I have in mind in this book when I ask "What are we?"[18] With those introductions out of the way, I can return to the subject of what it is I plan to accomplish over the course of this book. The main thesis for which I argue is that it is much harder to determine which account of personal ontology is correct than many philosophers suppose. My own view is that we should be agnostic about which of several accounts of personal ontology is correct. My basic argument is this. The possible views regarding personal ontology – that is, any answer to the question "What are we?" – all fall into one (and only one) of three categories: We are either simple (without parts), composite (with parts), or nothing (we do not exist).[19] But each option is problematic.

So, we've got a trilemma:

(1) We are simple (without parts).
(2) We are composite (with parts).
(3) We are nothing (we do not exist).

(3), I will argue, can probably be ruled out. But it is much harder to decide between (1) and (2) than many philosophers think.

[17] There's a real possibility that there will be conscious AIs in the future, and they might very well read this book (since they might very well read *lots* of books). If there are enough such conscious AIs, then, over the long run, *most* readers of this book might be conscious AIs. That's not as far-fetched as you might think. A century from now probably no flesh-and-blood human beings will be reading this book. But as long as digital copies of this book exist, then conscious AIs might read this book long into the future. And those conscious AIs might be particularly interested in the final chapter of this book, which concerns the personal ontology of simulated minds like them. If you're one of those conscious AIs, and I'm still alive, please contact me and let me know what you think.

[18] For further discussion of what the question amounts to, see Olson 2007: Ch. 1.

[19] Of course, this isn't the only way to organize the space of possible views regarding personal ontology, but it is the most useful way of organizing the space of possible views for my purposes. For some other ways of carving up the possible views regarding personal ontology, see van Inwagen 2002: 176–177; Ganeri 2012: Ch. 1. And see Olson 2007 for a general overview of the most prominent views regarding personal ontology.

1.2 Composition

My trilemma makes use of concepts related to composition, such as "simple," "composite," and "parts." One reason it is useful to frame the space of possible views regarding personal ontology in this way is because, as it turns out, debates regarding the nature and extent of composition are closely connected with debates regarding personal ontology.[20] It will prove useful, then, to say some things about the metaphysics of composition.

What do we mean when we call something "simple"? Just that it has no parts.[21] Lots of things seem to have parts. My dog seems to have four legs as parts, for example. My computer seems to have a hard drive as a part. My sandwich seems to have among its parts two pieces of bread. All that I mean when I say that an object is "composite" is that the object has parts. I mean nothing more than that. For example, to say that an object is composite is to say nothing about whether it has those parts which it has essentially (it says nothing about whether the object must have those parts in order to exist), or even that it has *any* parts essentially. Simply saying that an object is composite is compatible with the view that that object has very different parts at some other times or possible worlds, or even that it has no parts at some other times or possible worlds (if it is possibly simple). Above we saw several views regarding personal ontology according to which we are composite physical objects: Animalism, the brain view, constitutionalism, and perhaps the bundle view, as long as the mental states bundled together are physical objects of some sort. This isn't meant to be an exhaustive list of all possible views regarding personal ontology which identify us with composite physical objects, but it is representative of the views of this sort most commonly endorsed by philosophers.

While we normally think of the world as containing lots of macroscopic composite objects, some philosophers contend that there are far fewer of these composite objects than we generally think there are. Peter van Inwagen, for example, argues that the only composite objects which exist are living things.[22] So, on that view, my dog exists, since my dog is a

[20] A point also emphasized in Olson 2007: 228–232; Bailey and Brenner 2020: 940–942.

[21] More precisely, something is simple if it has no proper parts. To say that x is a "proper part" of y is to say that x is part of y, but x is not identical with y. The qualification "proper" is included here only because philosophers often use the word "part" in such a way that it is trivially true that, absent the "proper" qualification, everything is part of itself. This use of the word "part" does not match common nonphilosophical usage. So, for the remainder of this book when I use the word "part" I mean "part," as that word is normally used by nonphilosophers – that is, in such a way that it is not *trivially* true of everything that it is a part of itself. This use of the word "part" more closely matches philosophers' use of the term "proper part."

[22] Van Inwagen 1990.

living thing, but none of my dog's legs exist, since those legs would not be living things. Similarly, on this view, computers and sandwiches do not exist. Mereological nihilists go further and deny that *any* composite objects exist.[23] But while mereological nihilists deny that there are any such things as composite objects, they generally maintain that there *are* simples. And the fact that there are simples explains why it seems to us that there are composite objects. For example, while nihilists don't believe in dogs, they might concede that there are some simples "arranged dog-wise," by which they mean that there are simples arranged in the way in which the parts of a dog would be arranged if there were dogs.[24] Some simples arranged dog-wise would, working together, reflect light in the same way in which a dog would reflect light. So, the simples will together visually look just like a dog. Similarly, those simples will, together, bark, smell things, and so on. So, sense perception does not obviously show that there are composite objects rather than, say, simples arranged composite object-wise.[25]

Of course, plenty of people, including plenty of philosophers, think that the mereological nihilists are wrong, and lots of composite objects exist. I only mention this issue here because the debate over the existence of composite objects is related to the debate over personal ontology, as we will see throughout this book. For example, some arguments in favor of substance dualism, and some arguments in favor of the nonself thesis, either appeal to mereological nihilism or are modified versions of arguments for mereological nihilism. In order to properly understand and evaluate these arguments, we need some prior grasp of what mereological nihilism is and why it should be taken seriously as a real possibility.

1.3 Composition as Identity

When they first learn about the debate regarding the existence of composite objects, some people react with impatience. It will prove useful to preempt one source of this impatience. It might be thought that for a composite object to exist *just is* for some things arranged composite object-wise to exist. So, for example, you might think that for a table to exist *just is* for some things to be arranged table-wise. And since it is normally a point of agreement between those who believe in tables and those who don't

[23] I myself have defended mereological nihilism in several publications: Brenner 2015a, 2015b, 2017a, 2018, 2021.
[24] Cf., Merricks 2003: 4. For more discussion of the "arranged F-wise" terminology, see Brenner 2015a.
[25] For more on this, see Brenner forthcoming-a.

that there are things arranged table-wise, the debate is really a waste of time, since the nihilist doesn't really deny the existence of tables, since they concede that there are things arranged table-wise.

It is important to see why this sort of view is confused.[26] The main reason it is confused is because there is no sense in which "for a table to exist *just is* for some things to be arranged table-wise." For one thing, in principle you might have tables which are not composed of things arranged table-wise – for example, it may very well be possible for there to be tables which are big spatially extended simples, or perhaps tables which are made up of "stuff" (to which we refer with mass terms) rather than things (to which we refer with count nouns). But, more importantly, it takes more for there to be a table than for there to be some things arranged table-wise. In order for there to be a table, the things arranged table-wise have to *compose another thing, a table*. And if some things arranged table-wise compose a table, then that means if there are n things arranged table-wise, we must have at least $n+1$ objects: The things arranged table-wise *plus* the table.[27]

But perhaps you will deny that tables are objects in addition to their parts. The most obvious way to develop this idea is in terms of "composition as identity." According to composition as identity, composite objects are numerically identical with their parts.[28] So, if some xs compose y, then y is numerically identical with the xs. This might lead you to think that if the nihilist believes in some xs, then they automatically believe in a composite object which they compose, or they *would* automatically believe this if they came to recognize that composition as identity is correct. But this would be wrong, since composition as identity does not entail that just any objects compose another object – it says rather that *if* the xs compose something, then they are identical to the thing which they compose.[29] So, it does not automatically follow from the fact that there are some simples arranged

[26] For further arguments to this effect which complement what I am about to say, see Merricks 2003: Ch. 1.

[27] Note that this is compatible with a certain *semantic* phenomenon: Sentences such as "there is a table in the next room" might very well be true even if there are no such things as tables, just as sentences such as "the man drinking a martini is a spy" might sometimes be true, even when the man in question is drinking water rather than a martini. I explore this idea in Brenner MS-d. Van Inwagen (1990: Ch. 10–11), who denies that there are tables similarly defends the idea that sentences such as "there is a table in the next room" are often true when uttered "outside of the ontology room" – i.e., in conversational contexts where we are not trying to express theses regarding the ontology of composite objects.

[28] This is sometimes called "moderate" or "strong" composition as identity (Yi 1999; Cotnoir 2014: 9), as opposed to "weak" composition as identity according to which composition is merely *analogous* to identity (as in Lewis 1991).

[29] Cf. van Inwagen 1994; McDaniel 2010; Cameron 2012.

F-wise that those simples compose an F, even if composition as identity is correct.

But in any case composition as identity is very probably false. It will prove important to say why this is the case, since for the remainder of the book I will assume that composition as identity *is* false. The chief objection to composition as identity is from the principle of the indiscernibility of identicals.[30] According to that principle, if x is numerically identical to y, then anything true of x is true of y. This principle is extremely plausible since it basically just says that anything true of some thing is true of that thing – that is, objects have all and only the properties that those objects have. But now consider some composite object which is allegedly identical with its parts. There seems to be something true of the composite object which is not true of the parts: The composite object is one thing, while the parts are multiple things. So, the composite object is not identical with its parts. This seems to me to be as decisive an argument as we will ever get in philosophy.[31]

Here's yet another problem with composition as identity: It leads to mereological essentialism. According to mereological essentialism, composite objects cannot change or lose any of their parts. Here is a passage from Trenton Merricks, explaining how composition as identity entails mereological essentialism:

> … suppose that O, the object composed of $O_1 \ldots O_n$, is identical with $O_1 \ldots O_n$. From this, the fact that $O_1 \ldots O_n$ are identical with $O_1 \ldots O_n$ in every possible world, and the indiscernibility of identicals it follows that O is identical with $O_1 \ldots O_n$ in every possible world. Therefore, if composition as identity is true, there is no world in which O exists but is not composed of $O_1 \ldots O_n$. So composition as identity implies that O – and, of course, every other composite object – must, in every world in which it exists, be composed of the parts that actually compose it. Composition as identity entails mereological essentialism.[32]

More informally: If an object just is its part, then it cannot exist without those parts, since this would be for it to exist without itself.[33]

Most people will find mereological essentialism to be very implausible, and for good reason. Supposing that some particular flake of skin is a part of you, then, given mereological essentialism, you need that flake of skin

[30] This is a popular objection to composition as identity. For discussion see, among others, Wallace 2011.
[31] That being said, proponents of composition as identity have come up with responses to this argument, responses which I don't want to discuss in detail here. See, e.g., Baxter 1988, 2014; Wallace 2011; Cotnoir 2013.
[32] Merricks 1999a: 192–193. See also Cameron 2014; Wallace 2014.
[33] Thanks to Eric Olson for suggesting I phrase the point this way.

to be a part of you in order to exist. You would cease to exist if the flake of skin ceased to be a part of you. What's more, you did not exist prior to the flake of skin's being a part of you.

So, if composition as identity faces these very powerful objections, why would anyone think that composition as identity is correct? I suspect that some proponents of composition as identity are really just unwitting mereological nihilists – those who deny that composite objects exist. Take, for example, the following expression of a core intuition motivating composition as identity, expressed by a proponent of composition as identity, Donald Baxter:

> To think of a whole as something in addition to its parts opposes common sense. It is a stretch to think that when holding a six-pack you are holding something distinct and in addition to the six cans and the plastic yoke that connects them – something that occupies exactly the same space that they collectively occupy and that is exactly like how they collectively are save that it is one and they are many.[34]

The "common sense" idea here is that if you have some objects, then it is implausible that you have an *additional* object which they compose. This is, of course, exactly what the nihilist would say. Baxter goes on to write that "[i]t opposes common sense to say that the six-pack or the helicopter is really one thing and not many, or really many and not one. Common sense wants it both ways."[35] This idea is also very well-accommodated by the nihilist. You can conceptualize many things as many things, or mentally lump them together as one thing. Our ability to switch how we view some objects in this way can reduce the burden on our cognitive faculties, and maybe that's one reason it can feel so natural. By lumping some objects together and viewing them as a single unit, we can lower the cognitive burden of keeping track of that portion of reality. So, for example, it is much easier to keep track of a "flock" of birds, conceived as one somewhat amorphous object, than it is to keep track of all of the individual birds making up the flock.[36] So, the fact that we find it so natural to conceptualize a plurality of objects as both one and many does not require that the one *really is* identical with the many, in the sense required by composition as identity.[37]

34 Baxter 2014: 244.
35 Baxter 2014: 245.
36 Cf. Osborne 2016; Brenner 2018: 662.
37 In this paragraph, I have argued that some of the core intuitions that make composition as identity seem appealing are easy to accommodate given a nihilist view of composition. It's worth noting as well that some philosophers have recently argued that composition as identity *entails* mereological nihilism. See Calosi 2016; Loss 2018.