POLITICS & POETICS
A Journal for Humane Philosophy

Supplement

Transhumanism and the Metaphysics of Human Persons

Part of Issue on

The Human Difference
(2017 - 2018)

Transhumanism and the Metaphysics of Human Persons

Ralph Weir
University of Cambridge

Introduction

Transhumanism has gained attention in mainstream academia and the media at an impressive speed.[1] It is supported by thinkers of considerable distinction. Its general impetus, to use technology to benefit humans as much as possible, is obviously appealing. Despite this, a very large number of people respond to transhumanism with scepticism, aversion and even disdain. This is puzzling. How could a movement that has so much going for it invoke such a negative reaction? This paper proposes an answer to this question. I propose that transhumanism invokes aversion, at least partly because it appears to promote practical judgements that do not satisfy normal standards for practical deliberation. These judgements fail to satisfy normal standards for practical deliberation because they depend on unreliable metaphysical presuppositions.

This paper is especially interested in the *risks* associated with transhumanism. When people discuss the risks of transhumanist technologies they often talk about the potential effects of these technologies on socioeconomic equality, or on the meaningfulness of our lives. For example, some worry that widespread enhancements will result in a two-tier society, with powerful enhanced humans oppressing powerless 'naturals'. Others fear that a society in which we did not have to face the adversity of human frailty and mortality would be void of meaning. It is characteristic of these risks that they presuppose the *short-term success* of transhumanist technologies. The worry here is not that

---

transhumanist technologies will not perform their immediate functions, but that they will have undesirable long term effects.

By contrast, the risks I focus on concern whether transhumanist technologies will *work in the first place*. I have in mind, in particular, those radical enhancements whose viability is dependent on deep issues about the metaphysics of human persons. The most prominent examples are those technologies that involve largescale replacement of the human person with artificial components, or the wholesale transference (or 'uploading') of the mind to a computer. The success of these technologies is dependent on several broadly naturalistic assumptions about the metaphysics of human persons. I argue that transhumanists exhibit unwarranted confidence in these assumptions.

In the next section I set out the epistemological basis of my argument. Following this in section two I clarify what I understand by 'transhumanism' and note three 'naturalistic' metaphysical theses on which the probable success of transhumanist technologies depends. These are physicalism, computationalism, and the continuity theory of identity. In section three I advance five arguments against one or more of these naturalistic presuppositions. In section four I explain why these arguments should be taken seriously, even by those who, on balance, reject their conclusions. In section five I explain why this substantiates the thesis that our epistemic situation does not warrant a high degree of confidence that the immediate effects of transhumanist technologies will be desirable, rather than extremely undesirable. Finally, in section six I consider four responses that transhumanists might make.

Bayesian Epistemology

Sometimes the aversion people exhibit towards transhumanism is based in a wider disagreement. For example, a committed theist might take the view that transhumanist enhancements represent a hubristic attempt to improve on God's plan. Such objections might be correct. But in order to convince transhumanists of them it would be necessary to resolve long-standing disputes, such as the contention between atheists and theists, that predate the question of transhumanism. By contrast the objection raised in this paper is intended to challenge transhumanists on their own terms. This objection rests on two claims, one epistemological and one epistemic.

Traditional epistemology works with concepts of justified belief and knowledge. Discussions of transhumanism often proceed in this way. For example, the question is raised whether a particular transhumanist

technology, such as uploading, is viable. In answer, a specialist responds by offering a case for *believing* that the technology will work. If the case is compelling then we will come away with a justified belief that, given what we know about neurology, computer science, and similar, uploading is in principle a technological possibility. Christian Klaes' paper in Göcke and Meier-Hamidi (2018) takes something like this approach.

There is nothing wrong with proceeding in this way when what we are interested in what to *believe* about transhumanist technologies. But frequently, our interest in transhumanism is not merely theoretical but *practical*: we are not just interested in what to believe about transhumanist technologies, but in whether we should develop such technologies, build a society that esteems them, or to seek to undergo enhancement ourselves. Where our interest is practical the procedure just outlined is not appropriate. This is because it has been convincingly argued by epistemologists such as Kaplan (1983) that for practical deliberation, the traditional notions of justified belief and knowledge are much less relevant than Bayesian notions of degrees of confidence and probability. And crucially, one might well believe a thesis p, without having a degree of confidence sufficient to warrant *acting on p*.

My epistemological thesis, then, is just the Bayesian claim that practical deliberation should be done in terms of degrees of confidence, not justified belief. Bayesianism is not unfamiliar to transhumanists. Eliezer Yudkowsky, for instance, is a well-known proponent of both. But combined with my second thesis, it has an bioconservative consequence. My second thesis is that our epistemic situation does not warrant a high degree of confidence that the immediate effects of transhumanist technologies will be desirable, rather than extremely undesirable. The success of these technologies depends on strong presuppositions about the metaphysics of human persons. Even if we *believe*, on balance, that these presuppositions are true, our understanding of the metaphysical issues is far too rudimentary to accord them a high degree of credence. This situation is not likely to change soon. If these theses are true there are good grounds to think that our practical judgments concerning transhumanist technologies should, for the foreseeable future, be bioconservative ones. Before defending this claim it is necessary to outline what I take the metaphysical presuppositions of transhumanism to be.

Transhumanist Metaphysics
Transhumanism is a nebulous phenomenon with more and less serious poles. But I take it that transhumanism is at least a movement that

purports to make informed, rational, recommendations for the investigation and ultimate use of technological enhancements of human persons: that is, in addition to *theoretical* claims about how technologies *will* develop and be used, transhumanism makes *practical* claims about how these technologies *should* be developed and used. I also assume that these recommendations should live up to normal standards for practical deliberation, just like recommendations for health policy or traffic legislation.

Where transhumanists differ from those making recommendations for health policy or traffic law is in the radical nature of some of the technologies involved. Of course, not all human enhancements endorsed by transhumanists are radical. They exist on a spectrum. At one end are wrist watches and spectacles. Obviously one does not have to be a transhumanist to endorse these. At the other end are procedures that involve largescale replacement of the human, including any mental parts thereof, with artificial components, or wholesale 'uploading' of the mind on to a computer—and the putative benefits these might open up, such as immortality, superintelligence and non-local consciousness. In what follows I am interested only in those forms of transhumanism that promote the technologies at this more radical end of the spectrum. I discuss briefly in section six whether this means that there are forms of transhumanism to which my argument does not apply.

Transhumanist technologies stand a chance of working only if they are not based on false metaphysical assumptions. Transhumanists frequently rely on substantive metaphysical presuppositions. These presuppositions can usually be classed as 'naturalist'. Naturalism in metaphysics is roughly the thesis that there exists nothing over and above the kinds of things described by the physical sciences of physics, chemistry and biology. Naturalism has antecedents that stretch back as far as the physical theories of the pre-Socratic philosophers, and more recently the mechanical philosophy of the seventeenth century. But it was only in the mid-twentieth century that metaphysical naturalism gained widespread acceptance amongst philosophers, as well as considerable adherence in wider culture, especially in the works of popular scientists.

Metaphysical naturalism, as I have defined it, sometimes goes by the name of 'physicalism'. Physicalism is intuitively hospitable to transhumanism. This is because it represents humans and other animate beings as, essentially, complex machines or 'moist robots'. This is good news for transhumanism because transhumanists aspire to upgrade or replace the natural (as opposed to artificial) components of humans with

manmade technologies. If physicalism is true then, whatever practical obstacles this aspiration may face, it looks in principle possible. If physicalism turned out to be false, on the other hand, this would raise serious questions about the viability transhumanist enhancements. Metaphysical naturalism, or 'physicalism', then, is the first naturalist metaphysical presupposition on which transhumanist technologies depend. Two further presuppositions are the computational theory of mind (computationalism) and the continuity theory of personal identity.[2] (See e.g. Kurzweil, 383).

Since the mid-twentieth century, proponents of naturalism have increasingly modelled minds on computers. In its classical form, advanced by McCulloch and Pitts (1943), the computational theory of mind holds that the mind is something like a Turing machine. Since the 1980s this view has been rivalled by the neural networks models defended by Rumelhart et al. (1986 and 1987) Either way, the computational theory entails that the mind is nothing over and above certain computational states. This theory is presupposed by 'uploading' as it is usually understood. The useful thing about computational states for uploading is that they can exist in the same way when implemented in different substrata. So, although the computational states making up your mind may *happen* to be implemented in your nervous system, there is no reason why they could not be implemented in silicon chips or other devises instead.

The continuity theory of personal identity is a response to the problem of identity over time. It is conspicuous that things change over time. This raises the question, what is it that makes some future thing *x* the same thing as *you*? Various answers have been advanced. On one view the fact that something is you is a fundamental property, over and above its other characteristics. On a second view, the idea that you retain your identity over time in the first place is an illusion. On a third view, the continuity theory, a future thing *x* is identical with you if and only if it is appropriately *continuous* with you. Kurzweil calls this view 'patternism' and uses the image of the ripple made in a stream as it passes a stone. Although the water of which the ripple is composed changes, a continuous pattern remains continuous throughout. The continuity theory of identity is useful for transhumanists because it suggests that if artificial components are gradually substituted for the natural

---

[2] These theories are naturalist not in the sense that they *entail* metaphysical naturalism, but in that they are consistent with, and partially motivated by metaphysical naturalism.

components of human persons, whilst preserving a continuous pattern, the identity of that person will be preserved also.

   Transhumanist technologies, then, tend to rely on at least three naturalistic metaphysical presuppositions: physicalism, computationalism, and the continuity theory of identity. These do not exhaust the metaphysical presuppositions of transhumanism. But they will be sufficient to illustrate the difficulties transhumanism faces. In the next section I present five arguments for non-naturalist theses that directly conflict with one or more of these presuppositions. Some of these arguments have considerable long-standing support amongst metaphysicians; others are more novel. All of them however are logically compelling and raise serious challenges for transhumanism. The take-home point is that there are numerous respects in which human persons do not seem to be the kind of things transhumanists tend to assume that we are.

Challenges to Transhumanist Metaphysics
A. The Standard Argument for Property Dualism
In the sense of 'naturalism' defined above, the best known anti-naturalist argument in contemporary philosophy is the standard argument for property dualism about consciousness (henceforth 'property dualism'). Property dualism is the thesis that in addition to physical entities there are nonphysical experiential (or 'phenomenal') properties. Examples includes the colour of a region in one's visual field, or the character of sensations like pain and pleasure. Classic defences of property dualism have been advanced by Frank Jackson (1982) David Chalmers (1996) and Jaegwon Kim (2005). As Chalmers (2010) notes, although arguments for property dualism differ in important ways, they tend to share something like the following structure:

1. The physical facts about the world do not a priori entail the existence of consciousness
2. If the physical facts do not a priori entail the existence of consciousness, then they do not necessitate the existence of consciousness
3. Therefore, the physical facts do not necessitate the existence of consciousness

This argument is logically valid: if the premises, 1 and 2, are true, then the conclusion, 3, must be true too. The conclusion says that the physical facts about the world do not 'necessitate' the existence of consciousness.

The claim that A 'necessitates' B is just the claim that it is impossible for A to exist or obtain in the absence of B. If the physical facts about the world do not necessitate the existence of consciousness, it follows that conscious states are something over and above physical things—they are something *extra* that still needs to be added once all the physical things are in place. So, if the premises of this argument are true then physicalism is false.

Premiss 1 can be elucidated in a number of ways. To say that A 'a priori entails' B is to say, roughly, that it is impossible to imagine A without B, or that if one knows that A exists or obtains one can rationally infer that B exists or obtains too. So, the idea of premiss 1 is that, even if you knew everything that the physical sciences could tell you about the world, and had the concept of consciousness, you still would not be in a position to infer that anything was conscious. This is often illustrated by using the idea of a philosophical 'zombie' or a 'zombie-world'. A zombie is a physical duplicate of an actual conscious person that is *not itself conscious*. A 'zombie-world' is a physical duplicate of the actual world where nothing is conscious. Premiss 1 can be understood as the claim that even with all the physical facts at our disposal, we could still imagine a zombie or a zombie world.

The same kind of idea is supported by a classic thought experiment advanced by Frank Jackson (1982). Suppose there is a scientist called Mary who is an expert on colour perception. Mary knows everything that physical science can tell us about the way humans perceive colour. But Mary has never seen colour herself—she has spent her life in a room where everything is black or white. Intuitively, when Mary first leaves the room, and sees the blue of the sky, and the red of a tomato, she learns something new about colour perception that her scientific knowledge did not give her. Before, Mary did not know *what it is like,* experientially, to perceive colours. She only learns this when she perceives colours for herself. If this is right, then the properties involved in Mary's conscious experience seem to be beyond those whose existence one can infer from the physical facts.

Thought experiments like philosophical zombies and the story of Mary are highly compelling. As a result, premiss 1 of the above argument is relatively uncontroversial. Even leading opponents of dualism tend to accept that it is true. Premiss 2 is more controversial. According to premiss 2 if the physical facts about the world do not a priori entail the existence of consciousness then they do not necessitate the existence of consciousness. Again, 'A necessitates B' just means that it is impossible

for A to exist or obtain without B. So premiss 2 is saying that if the physical facts about the world do not a priori entail the existence of consciousness then all the physical facts could obtain *without there being any consciousness*. It is this that makes it look like conscious states are something over and above physical things.

Premiss 2 is an instance of a much-disputed principle. According to this principle, necessitation has to be accompanied by a priori entailment. The appeal of this principle can be illustrated by an example drawn from a well-known scene in Lewis Carrol's *Alice in Wonderland*. At one point in Alice's adventures she encounters a Cheshire cat which is described as 'grinning from ear to ear'. After a few exchanges the cat slowly vanishes, leaving behind only its grin 'which remained some time after the rest of it had gone'. In her astonishment Alice remarks `Well! I've often seen a cat without a grin… but a grin without a cat! It's the most curious thing I ever saw in my life!'. (Carrol, 93)

This passage is amusing because Carrol asks us to accept an impossibility. We know, of course, that a grin cannot really exist on its own, without the thing that was grinning. This raises the question, *how* do we know that the cat's grin cannot exist on its own without the cat, whereas its tail, for instance, could? The most obvious answer is that we know this because necessitation has to be accompanied by a priori entailment. If we try to imagine a grin without a bearer we cannot do so, and so we infer correctly that a grin necessitates the existence of a bearer. But if we try to imagine a tail without an owner, we can, and so we infer correctly that a tail does not necessitate the existence of its owner. Our concepts of these things seem to be enough to tell us what they do or do not necessitate. Since this kind of reasoning works here, it seems it should work in the case of physical things and conscious experience too.

The principle instantiated in premiss 2, then, is intuitively appeling. Its validity is nonetheless subject to an ongoing debate which cannot be discussed in detail here. In short, there seems to exist a class of exceptions to the principle. These were identified by Kripke (1980). This has led many to conclude that the principle should be rejected despite its initial plausibility. On the other hand, there are grounds for thinking that the kind of exceptions to the principle that Kripke identifies *cannot include* the relation between consciousness and physical things relevant to premiss 2. And so, as Kripke himself argues, even if the principle fails in some cases it seems that it should be accepted in arguments for dualism after all. The debate continues and there is no sign that it will reach a consensus

soon.[3] Until this happens the argument for property dualism cannot be considered decisive. But neither can it be rejected out of hand.

The argument presented in this section poses a serious challenge to physicalism. As a result, it poses a serious *prima facie* problem for transhumanism. After all, if some kinds of mental entities are necessarily nonphysical then physical inventions cannot be substituted for them. Moreover, if the mental is something over and above the physical then it is, *a fortiori*, something over and above the computational states implemented in physical things. It follows that however good a model Turing machines or neural networks provide for mental states, the computational theory of mind is false. So, if the standard argument for property dualism is successful then transhumanist technologies that rely on naturalistic metaphysics require considerable rethinking.

B. Argument from Property Dualism to Substance Dualism
In one form or another, the above argument had become very popular over recent decades. Property dualism is now recognised as one of the leading contemporary theories of mind. One reason people are drawn to property dualism is that it is still considered *relatively* 'naturalist' when compared to more extreme positions. Its leading defenders have tried to persuade their readers that the discovery of nonphysical experiential properties would be analogous to a normal scientific advance, like the discovery of electromagnetic force. It would be an incremental addition to our ontology, not a radical revision. They have thus tried to distance their position form the more extreme 'substance dualism' associated with thinkers like Plato, St Augustine, and especially Descartes.

The important difference between *property* dualism and *substance* dualism is that properties cannot exist on their own whereas substances can: properties are *metaphysically incomplete*. (c.f. Williams, 7) This means that like the grin of Carrol's cat, properties necessitate the existence of something further beyond themselves. By contrast a 'substance' in the sense that is relevant here, is something that is *metaphysically complete*. A substance is something that *can* exist on its own.[4] Although property dualists think that there are nonphysical mental *properties* they think that these always belong to *physical substances*. Substance dualists, on the other hand, think that there are nonphysical mental substances—mental things that could, in principle, exist without anything physical.

---

[3] Much of this debate is traced in García-Carpintero and Macià (2006).
[4] Descartes gives this definition of substances (CSMK II 159-210).

The most familiar example of a mental substance is the soul. As it is usually conceived, the soul is a nonphysical thing that can exist without physical things like the body. Because substance dualism can accommodate souls it is congenial to religious theses about the afterlife and to paranormal phenomena such as ghosts. As a result, substance dualism is widely seen as being significantly more anti-naturalist than property dualism. It might be hoped, on these grounds, that property dualism will be less of a threat to transhumanist technologies than substance dualism would be. Perhaps minor revisions to the way transhumanist technologies are understood could accommodate nonphysical properties. If so then the argument of the last subsection may not be so bad for transhumanists.

This is probably hopeful thinking. Once there are nonphysical properties that are fundamentally different to physical properties the viability of transhumanist technologies as usually understood seems highly questionable. But even if this were not so, the challenges for transhumanism would not stop there. For, whilst the popular view is that it is possible to rest with property dualism and avoid the more thoroughgoing non-naturalism of substance dualism, it is not clear that this is really a defensible position. This is because there exists a compelling argument that once property dualism has been accepted, substance dualism has to be accepted too. The simplest way to put this argument is as follows:

1. The facts about consciousness do not a priori entail the existence of anything physical
2. If the facts about consciousness do not a priori entail the existence of anything physical, then they do not necessitate the existence of anything physical
3. Therefore, consciousness does not necessitate the existence of anything physical

We already know that if the argument for property dualism is successful then consciousness is not physical. According to this argument consciousness does not necessitate the existence of anything physical either. This means that there are nonphysical things that can exist on their own without physical things. It follows that there are nonphysical substances. The argument is logically valid, so if its premisses are true the conclusion must be true too.

The first premiss of this argument is roughly the claim that, even if you knew everything about people's conscious experience, you still would

not be in a position to infer that there is anything physical. This seems highly plausible because conscious experience does not seem to entail anything beyond itself. In his classic defence of property dualism, Chalmers himself (1996, 75) endorses something like this thesis when he acknowledges that the facts about the external world do not 'supervene logically' on the facts about experience. The second premiss is simply an instance of the principle that necessitation has to be accompanied by a priori entailment. As we have seen, this principle is controversial. But if it is accepted in the standard argument for property dualism, then consistency demands that it be accepted here too.

It seems then that if the argument for property dualism is successful, then there is good reason to suppose that the argument for substance dualism will be successful too. The central point of this argument can be put in another way. Properties are metaphysically incomplete because they need something further in order to exist. For example, the pitch of a note cannot exist without a volume and a timbre. Likewise, the size of an object cannot exist without a shape. So, if consciousness only involves nonphysical *properties*, there should be some physical thing that stands in the same conceptual relation to conscious experience as shape does to size or timbre to pitch. But there seems to be no such thing. Rather, conscious states seem to be *conceptually independent* of anything physical.

Is there a way for someone who accepts the standard argument for property dualism to reject this argument? I think the best strategy would be to claim that consciousness a priori entails the existence of a *subject* of conscious states, and that this subject is a physical thing. This manoeuvre is not obviously workable however. For it is plausible that subjecthood is just a second order property that a substance has in virtue of having experiential properties. If so then subjecthood is not some *additional* property that might turn out to be physical. It might be argued against this that we have a substantive notion of a subject in its own right, perhaps as something like Descartes 'thinking thing' or Husserl's 'transcendental ego'. But in that case, it is doubtful that it will be something the physical world can provide.

If this is correct then the grounds for thinking that property dualism obtains are also grounds for thinking that substance dualism obtains. This may pose an even greater challenge to transhumanism than property dualism alone. For even *if* nonphysical properties *do* constitute only a minor addition to the naturalistic ontology that transhumanists typically presuppose, nonphysical substances obviously constitute a major, radical

addition. To take one example, if conscious subjects are nonphysical substances, then transhumanist strategies for evading death may be completely misguided. For all they will achieve is the prolongation of certain physical mechanisms whose continuation is neither necessary nor as far as we know sufficient for the continued existence of the subject.[5]

C. The Argument from Being You

The previous two subsections present logically compelling arguments against the naturalist presuppositions of transhumanist technologies. Both are concerned with the nature of consciousness. Perhaps a transhumanist will respond that consciousness is not well enough understood for us to put serious weight on these arguments. After all, it is well known that consciousness is an unusually mysterious phenomenon, and one that physical science hasn't yet come to grips with. But, they may urge, it is only a matter of time until this changes, and then the way for radical enhancements will be clear.

This is not correct. For there also exist logically compelling arguments against naturalism about human persons that focus on things other than consciousness. One such argument, advanced in some form by Stephen Priest (2000) and Benedikt Göcke (2012) concerns personal identity.[6] This argument is usefully introduced by considering some of the differences between *you* and other people. Here is the situation as presented by Priest:

> You look out of your own eyes, but you look at or into other people's eyes. You have never seen your own face, nor the back of your own head. In the case of just one human body (the one called your own) you feel yourself to be wholly or largely co-extensive with it. Perhaps you are inside your body or perhaps you are your body, looking out of it. In just *this* body but nobody else's you experience sensations and thoughts. The rest of the world seems to be arranged around you, with your body at its centre. (Priest 2000, vii)

These phenomenological and physical facts are 'symptomatic of being oneself'. Noticing these facts allows one to understand a group of related questions: what is it for something to be you; why is something you; and what have you claimed about something, when you have claimed that it is you? Priest does not offer an answer to these questions. But he does argue that being you is something over and above all of the other properties, both experiential and physical, in terms of which we usually

---

[5] For alternative arguments from property dualism to substance dualism see Schneider (2012) and Zimmerman (2010).

[6] Though he does not invest much interest in it, Chalmers (1996, 85) also touches on something like this argument.

characterise a person. (152-3) Adapting the argument for present purposes we can state the following grounds for thinking that this is so:

1. The physical and experiential facts about you do not a priori entail that you are you
2. If the physical and experiential facts about you do not a priori entail that you are you, then they do not necessitate that you are you
3. Therefore the physical and experiential facts about you do not necessitate that you are you

Unlike those above, this argument does not aim to show that naturalistic metaphysics cannot accommodate consciousness. On the contrary, according to this argument, even once all the physical facts *and* the facts about your conscious experience have been accommodated, there is *still* something further that these do not necessitate: the property of being you. If the argument is successful it follows that *being you* is a further nonphysical property.

The argument is logically valid. Premiss 2 depends on the principle that necessitation must be accompanied by a priori entailment. Since I have discussed this principle above I will not go over it again here, except to say that for reasons that cannot be entered into now, I think it unlikely that this premiss will fall into the class of exceptions to the principle identified by Kripke. If premiss 2 is accepted, then the success of the argument depends on premiss 1. Premiss 1 is roughly the claim that someone could know everything there is to know, both physically and experientially, about you, without being in a position to infer that the object of their knowledge *is you*.

This can be supported by thought experiments involving doppelgangers. Although modern science is not capable of it, we can imagine a scenario in which you are physically and mentally duplicated, so that after the process, in addition to you, there is a second human being with the same physical properties and the same experiential properties as you. *Ex hypothesi* only one of these beings is identical to you. But since they share all of their physical and experiential properties it seems that nothing we could know about these properties would allow us to infer that one doppelganger is you and that the other isn't. So, this thought experiment seems to show that the physical and experiential facts about you do not a priori entail that you are you.

If the premisses of this argument are true, then it is not just consciousness that poses a challenge to transhumanist metaphysics. For

even if the naturalistic picture that transhumanists tend to presuppose can accommodate consciousness, it still won't be able to accommodate personal identity. This poses a further threat to transhumanist enhancements. If *being you* is a nonphysical property, then it is *a fortiori* something over and above any computational state implemented in your nervous system, and any pattern of continuity that your physical (or experiential) properties exhibit. So, the conclusion of this argument undermines enhancements that rely on physicalism, computationalism, or the continuity theory of identity.

The argument of this subsection is closely related to a well-known argument about personal identity over time advanced by Derek Parfit (216-227). Parfit develops problems for the continuity theory of identity, beginning with the observation that it seems to be consistent with continuity relations that *two or more* entities could be continuous with the same human person. For example, suppose we scan the computational state of your mind before you die, and upload it to a computer. According to many transhumanists that upload would satisfy a continuity relation that makes it identical to you. But of course, we could upload it to several computers.

These uploads are not identical to one another. And it is a law of logic that identity is a transitive relation. So, it follows that the uploads cannot be identical to you. If this is correct then the continuity relation between you and the uploaded computational state is not sufficient to preserve identity. Although Parfit uses his argument to challenge the notion of personal identity generally, it can also be used to support the thesis that being you is something over and above your physical and experiential properties. The relationship between Parfit's argument and transhumanism has been discussed at greater length by Susan Schneider (2016).

D. Argument from the Unity of Consciousness

The problems posed by personal identity and consciousness to naturalist metaphysics have received a relatively large amount of attention in recent years. A family of arguments that has received less attention recently concerns a particular characteristic of conscious experience: its *unity*. The idea that the unity of consciousness may be at odds with naturalistic metaphysics has its roots in classic treatments such as Descartes' *Meditations* and Kant's *Critique of Pure Reason*. More recent arguments have been put forward by Hasker (1995) and Zimmerman (2010). Here I briefly present one such argument.

The term 'unity of consciousness' has been used to refer to numerous phenomena. One example is 'binding' or 'object unity'. This occurs when a subject experiences an object as uniting distinct kinds of perceptual feature such as shape and colour. (Tye 11-12) A second example is spatial unity. This occurs when a subject's conscious states are experienced as belonging to a common space. (Tye 12; Bayne and Chalmers 25-6) A third example is the fact that our conscious experiences seem 'homogenous' as opposed to 'grainy'. (Lockwood 1993)

In this section I am interested in what is described as 'subject unity'. 'Subject unity' refers to the relationship between conscious states that are simultaneously undergone by the same subject. Subject unity also involves 'phenomenal unity'.[7] This term refers to the way different conscious states are 'experienced together' when e.g. one hears a piece of music whilst observing scenery. (Bayne 10)

For example, compare the following two scenarios. In the first scenario *I* hear music whilst *someone else* observes a landscape. In the second scenario *I* hear music and *I simultaneously* observe a landscape. The relationship between the two experiences in the first scenario is strikingly different from the relationship between the two experiences in the second scenario. In the first scenario, the two experiences are completely separate. In the second scenario, they seem to be two parts of one overall experience. (Bayne and Chalmers 23-7; Bayne 11)

There are two things that are interesting about the unity of consciousness illustrated by this example. The first interesting thing is that this unity seems to be a strict, all-or-nothing matter. We cannot imagine half a conscious subject as we can imagine half a body. And there is nothing vague or indeterminate about whether a thought or feeling belongs to my conscious experience, or to yours. On the contrary, as William James (226) puts it, 'the breaches between such thoughts are the most absolute breaches in nature'. The second interesting thing is that this strict, all-or-nothing unity occurs at the *human scale*. In usual cases, at least, there seems to be exactly one unified field of conscious experience per human being (rather than per quark, for instance, or per universe).

In combination, these two characteristics of the unity of consciousness pose a new problem for the naturalistic metaphysics presupposed by transhumanism. On a naturalistic view, conscious subjects are nothing over and above the kinds of things described by the physical sciences. For

---

[7] As some people use the terms it is possible to have subject unity without phenomenal unity. But I know no example of subject unity without phenomenal unity except where 'subject' is used to refer to something other than the thing that experiences.

example, conscious subjects might be identical to human organisms, nervous systems, or processes or states involving these. But these things seem to be *radically different in structure* to conscious subjects. This is because physical things at the human scale, at least, invariably seem to lack the kind of strict, all-or-nothing unity that conscious subjects have.

Human scale physical objects are like hills and valleys, in that they do not seem to have precise boundaries. There is no non-arbitrary answer to the question where a hill ends and where a valley begins. The same is true of human organisms, nervous systems, and processes and states involving these. These considerations suggest an argument like the following:

1. Conscious subjects have strict human-scale unity
2. Physical things do not have strict human-scale unity
3. Therefore, conscious subjects are not physical things

As before, the conclusion of this argument follows logically from its premisses. The first premiss seems at least highly plausible. Conscious subjects certainly seem to exhibit the kind of strict unity described. And this unity is certainly human-scale in the sense that there appears to be exactly one conscious subject per normal human being.

The second premiss also seems *prima facie* true. Take any recognisable human-scale physical thing, and you will find that it lacks the kind of strict unity that conscious subjects seem to have. Perhaps fundamental physical entities like fermions or bosons might have this kind of unity; and perhaps the physical universe as a whole does. So if there were one conscious subject per fermion, or one conscious subject per universe, the unity of consciousness would not pose a problem for naturalism. But the metaphysical structure of conscious subjects stands out conspicuously against the background of the human-scale physical world.

When we think in terms of conscious states, rather than subjects, we have a tendency to neglect difficulties of this kind. For example, the most prominent neurological criterion of consciousness is that advanced by Crick and Koch (1990). According to this account the neural states associated with consciousness are those that involve temporarily correlated firings as a result of coherent 'semi-synchronous' oscillations in the range of 40-70 Hz. The problem with this view is that there is no obvious way to count how many neural systems meet this description. It seems that there will be countless ways of carving up a region of nervous tissue that satisfies this criterion. As such, correlated neural firings do not seem like the right kind of things to account for the structure of conscious subjects.

Perhaps the most impressive rival to this account is the integrated information theory (IIT) defended by Tononi (2008). According to the IIT a system is conscious to the degree that it contains information that cannot be localized to the system's parts. Tononi also proposes that a system is only conscious if it is not fully contained by a larger system with greater integration of information. This manoeuvre is designed to make it plausible that there will usually be one appropriately integrated system per human being. But it is not clear this will work. For it seems that there will still be countless physical aggregates, differing at the atomic level, with an equal claim to being a realizer of any human-scale informational system. Furthermore, Tononi's theory has the counterintuitive consequence that many inanimate physical systems like photo-diodes are conscious.

If the argument of this subsection is correct, it once again follows that conscious subjects are something over and above the kinds of entities described by physical science. Again, this poses a problem for transhumanist technologies that presuppose physicalism or computationalism. Note also that unlike the previous arguments, this one does not depend on the controversial principle that necessitation must be accompanied by a priori entailment. Furthermore, even if this argument is not accepted, until a convincing physical correlate of the unity of consciousness *is found*, it is hard to imagine how transhumanist technologies could hope to guarantee the continued existence of a conscious subject.

E. Argument from Free Agency

We have seen that consciousness *tout court*, personal identity, and the unity of consciousness all pose challenges to the naturalistic metaphysics of transhumanism. The final challenge I present here concerns agency, in particular the kind of *free* agency that we usually assume is a necessary condition for moral responsibility. If naturalistic metaphysics cannot accommodate free agency, and transhumanist technologies are premised on naturalistic metaphysics, this raises the worry that transhumanist enhancements will divest us of free will.

The classic metaphysical debate in this area focuses on whether free will is compatible with physical determinism. Physical determinism is the thesis that at any moment in history, the past and the laws of nature *fully determine* everything that will happen in the future, including the behaviour of human beings. This idea was famously illustrated by Pierre-Simon Laplace (1902) who posits that if somebody knew the exact

momentum and location of every atom in the universe, then they could calculate their momentum and location for any future time by the laws of classical mechanics.

Many have felt that if physical determinism is true, then we cannot really have free will of the kind needed for moral responsibility. Peter van Inwagen (1983) formalises this idea in what is known as the 'Consequence Argument'. He summarises the argument as follows:

> If determinism is true, then our acts are the consequence of laws of nature and events in the remote past. But it's not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us. (56)

Van Inwagen's argument is intuitively compelling. Usually we think that people are not free to do the impossible, and cannot be held responsible for failing to do so. But if determinism is true then there is a fairly straightforward sense in which it is physically impossible for any person to act otherwise than they in fact do.

Physics has moved on since Laplace's time. On standard interpretations of quantum mechanics, determinism is not true. Rather, some events are undetermined—the past and the laws of nature leave open whether or not they will happen. Many, like Kane (1996) have appealed to quantum mechanical indeterminism in order to explain how free will might fit into the natural world. This strategy faces two major difficulties.

First, it is not clear that indeterminism at the very small quantum scale will result in non-negligible indeterminism at the scale of human action. Biological systems such as human organisms are purported to be too 'warm wet and noisy' to amplify quantum effects to the macroscopic scale. In recent years, this assumption has been challenged by evidence of quantum phenomena in biological contexts including plant photosynthesis, bird navigation, and micro-tubules in the brain. (Engel et al. 2007; Gauger et al. 2011; Hameroff and Penrose 2014) But the significance of these findings is not yet clear.

Even if quantum indeterminism is non-negligible at the human scale, however, there is a further problem for using it to account for free will. For although standard interpretations of quantum mechanics posit undetermined events, these occurrences are supposed to be *random*. And just as it is hard to accept that a free action could be determined by the distant past and the laws of nature, it is also hard to accept that a free action could be a random occurrence. If the only undetermined events are random ones, then it is natural to agree with Hobart:

> In proportion as [someone's action] is undetermined, it is just as if, his legs should suddenly spring up and carry him off where he did not prefer to go. Far from constituting freedom, that would mean, in the exact measure in which it took place, the loss of freedom. It would be an interference, and an utterly uncontrollable interference, with his power of acting as he prefers. (1939, 7)

These considerations have led to a variety of responses. Some like Hobart insist that free actions can be determined by the past and the laws of nature after all. Others like Kane hold that they can be undetermined quantum events. It will take much more work before any consensus is reached. But there are at least *prima facie* grounds for thinking that free actions can be neither deterministic, nor random events. This suggests that the following argument should be taken seriously:

1. Free actions cannot be deterministic or random
2. The events described by physical science are either deterministic, like those of classical mechanics, or random like some events in quantum mechanics
3. Therefore, free actions are not among the events described by physical science

Like the previous arguments in the section, this one is logically valid. Its first premiss is at least highly intuitive. Its second premiss reflects the current state of the natural sciences as they are usually understood. The conclusion constitutes a final challenge to the ontology typically presupposed by transhumanists. As adherents of metaphysical naturalism, transhumanists typically assume that there are no objects or events over and above those described by physical science. This argument suggests that this is false.

Of course, this is only the case if we assume that there *are* free actions in the first place. But this assumption is extremely widespread. It is presupposed by our legal systems, our understanding of history, and in our everyday lives. It is also presupposed by debates over the moral status of transhumanism. For if we do not have the kind of free will necessary for moral responsibility, transhumanism can be neither a praiseworthy nor a blameworthy enterprise in the first place. To respond to this argument by denying the existence of free will would therefore be a costly strategy for transhumanists.

Supposing we do have free will, the above argument raises the question what effect radical enhancements like uploading might have on our agency. The line of argument given here suggests that free actions are a kind of non-random undetermined event. Although we do not know of

any such events in the physical world, it is plausible that we are phenomenologically acquainted with them. If this picture of agency is correct, and if transhumanist technologies involve only the kind of events that physical science acquaints us with, it seems that these technologies cannot preserve free agency.

Non-Naturalism Cannot be Brushed Off

Transhumanist technologies rely on a number of naturalistic metaphysical presuppositions. The previous section has advanced five lines of argument against these presuppositions. All of the arguments are logically valid. In order to reject them it is necessary to reject one or more of the premisses. These are not the only arguments that I could have presented. They are a representative selection of some of the obstacles facing the kind of metaphysical presuppositions that transhumanists make.

It is likely that some transhumanists will respond with incredulity. 'Of course', they might say, 'there are some eccentrics who argue for fanciful theses like the immateriality of the self, or the irreducibility of agency. But don't serious, scientific thinkers reject these ideas as irresponsible speculation and wish fulfilment?' This is a widespread attitude to non-naturalist positions in metaphysics. In a longer essay it would be interesting to enter into some of the socio-cultural influences behind this attitude. I have argued elsewhere that naturalists are no less influenced by such influences than their non-naturalist opponents. But for present purposes I confine myself to the following three considerations.

First, it has long been a rhetorical strategy of naturalists to imply that non-naturalist views about human persons are only taken seriously by philosophers, theologians, and other non-scientists, who have yet to catch up with discoveries of modern science, especially neuroscience. This picture is misleading. In the mid-twentieth century, when metaphysical naturalism in philosophy was at its peak, many leading neuroscientists including Wilder Penfield, (1975) John Eccles, (1992) and the founder of modern neurophysiology C. S. Sherrington (1947) endorsed or gave weight to dualist views resembling those mentioned above.

Anecdotal evidence suggests that these thinkers are not exceptional nonconformers. Oliver Sacks (1990) reports that a large minority of specialists in neuroscience explicitly reject materialist theories. Presumably a greater number still considered the dispute between naturalist views and their alternatives undecided. During the week of the conference at which this paper was first delivered, the English

neuroscientist Raymond Tallis appeared on BBC Radio 4, criticising the naturalistic metaphysical presuppositions of transhumanism, which he describes as a 'travesty on the nature of consciousness'.[8]

There is little reason, therefore, to think that non-naturalist views about human persons are simply scientifically uninformed. Moreover, as is clear from the relevant literature, the methods of empirical science are not generally suited to resolving the metaphysical disputes in the first place. A second misconception is that non-naturalist views are no longer taken seriously amongst mainstream philosophers of mind. Again this impression is mistaken. Admittedly, there exist *some* philosophers who adhere to Daniel Dennett's 'apparently dogmatic rule that dualism is to be avoided at all costs'. (Dennett 1993, 5) But they are not in a noticeable majority. In fact, as Koons and Bealer (2010) point out, of major philosophers of mind active since the mid-twentieth century 'a majority, or something approaching a majority, either rejected materialism or had serious and specific doubts about its ultimate viability.'

The idea that opponents of naturalism are scientifically uninformed, or are a peripheral minority is therefore unfounded. In fact, if we want to appeal to scientific or philosophical authorities on the theory of mind regarding how seriously we should take the arguments presented in section three, the answer will be that we *should indeed* take them seriously. But it is worth emphasising that this is not a case in which we should need to appeal to authority in the first place. This is the third reason why non-naturalist views cannot simply be brushed off. The conclusions of the arguments presented in the previous section are logically entailed by their premises. As such, it falls upon *anyone* who accepts the minimal logical constraint of avoiding contradiction to respond to these arguments if they want to defend the viability of transhumanist technologies. Matters of logic hold regardless of what authorities in other disciplines have to say.

Transhumanism and Russian Roulette

Transhumanists cannot reasonably respond that the arguments advanced in section three do not need to be taken seriously. But they have another obvious line of response available. 'Although these arguments deserve serious attention,' they might say, 'on balance we don't accept their conclusions. In each case we think it is more plausible that one of the premisses is false, than it is that the conclusion is true. Therefore, we are

---

[8] <http://www.bbc.co.uk/programmes/b08ksc5h> accessed 31st July 2017.

still warranted in believing in the viability of technologies that depend on the naturalistic presuppositions that these arguments challenge'.

I think many transhumanists would be tempted to advance something like this response. After all, for every non-naturalist argument I have given there will be various naturalist counterarguments, and it is characteristic of metaphysics that these debates will not be drawn swiftly to a close. So, it might seem that whilst these arguments make *some of us* sceptical about the viability of radical human enhancements, on the basis of *their own* beliefs, transhumanists can still responsibly promote such technologies. To see what is wrong with this response it is necessary to return to the epistemological claim of section one. For it is here that the difference between the traditional epistemological notion of justified belief, and the Bayesian notion of degrees of confidence becomes important.

Practical deliberation should be done on the basis of one's degrees of confidence, not what one believes. This is because it is quite possible to have a belief that p without having a degree of confidence in p sufficient to warrant acting on p in all situations. This can be seen from the following example: suppose I have a revolver with one round in the cylinder. Mathematically, I should be about 83% confident that firing the revolver at my head will not result in my death. This might well be sufficient to warrant the *belief* that this course of action will not result in my death. But when deciding whether to *perform this action* this belief is not the relevant factor. Rather, what is important seems to be on the one hand my *degree of confidence* that this action will not kill me, and on the other hand the value I place on my life. Since I am very keen to avoid killing myself, I am probably under a rational obligation not to fire the gun, *despite the belief* that doing so will not have this result.

This reasoning can be transferred to the case of transhumanism. Although its leading proponents are typically familiar with Bayesianism, discussions of transhumanism are frequently carried out in terms of justified belief. For example, when an opponent questions whether the computational theory of mind is true, transhumanists marshal arguments aimed at showing that the theory should be believed. But as the revolver example makes clear, this is not the relevant factor when deciding whether to undergo uploading. For the result of uploading, if the computational theory of mind turns out to be false or incomplete, could be extremely undesirable. And as the arguments of section three show we have at least *some* serious grounds to doubt computationalism. If follows that *even if* we *believe* that the theory is true, we are unlikely to have a sufficient degree of *confidence* in its truth to make uploading rationally justifiable.

This judgement is based partly on the assumption that if the naturalistic presuppositions on which transhumanist technologies depend are false, their effects could be extremely harmful, and therefore the risks of undergoing them will be extremely high. It is difficult to make a precise evaluation of this risk, because our metaphysical understanding of the relevant topics—consciousness, identity, agency—is so rudimentary. But presumably the potential harms of transhumanist technologies include death; loss of freedom/agency; psychological fragmentation; and countless varieties of indefinite suffering. In these circumstances, it seems implausible that a proponent of transhumanism could be anything like confident enough in the thesis that radical enhancements will have desirable rather than extremely undesirable consequences to warrant the kind of promotion they accord these technologies.

Moderation and Mortality

I think that the line of argument advanced in this paper articulates a widely felt aversion to transhumanism. I also think that, pending *massive* advances in metaphysics, that aversion is probably justified. But I will briefly comment on three further responses a transhumanist might make to the argument developed here. The first is that transhumanists need not make the naturalist presuppositions that I have attributed to them. The second is that transhumanism is, or could be, a more moderate position than I have represented it as being. The third is that my argument fails to take account of the extreme undesirability of the situation we humans are already in.

I have said that transhumanist technologies rely on metaphysical naturalism. This might be questioned. For as Schneider points out, non-naturalist views like substance dualism do not *have* to conflict with transhumanism:

> It should be noted that although a number of bioconservatives seem to uphold the soul theory, the soul theory is not, in and of itself, an anti-enhancement position. For why can't one's soul or immaterial mind inhere in the same body even after radical enhancement? (2016, note 10)

So, it might be objected that the arguments of section three need not pose a challenge to transhumanism after all. For example, suppose conscious subjects are nonphysical substances. Clearly these substances stand in some kind of systematic relation to their physical bodies. Suppose those human bodies are gradually replaced by artificial components, or their computational states are uploaded. It is logically possible that after such

procedures the nonphysical substances will stand in the same systematic relation to the resultant transhuman bodies.

Although this is true, it does not pose a threat to the argument I have presented. For that argument does not rely on the claim that the success of transhumanist technologies is *incompatible* with non-naturalist positions like substance dualism. It only depends on the weaker claim that we have no right to be *confident* in their success, if non-naturalist positions are not ruled out. If a transhumanist could plausibly defend a mechanism that would *ensure* that radical enhancements will work, even if non-naturalist metaphysical theories turn out to be true, that would significantly increase the plausibility of transhumanism. But at present no such mechanism has been defended.

The second response is that transhumanism is, or could be, more moderate than I have represented it as being. There are two ways this response might be advanced. First, it might be pointed out that whilst I have treated radical enhancements like uploading as essential to transhumanism, some self-identified transhumanists express little interest in these technologies. So, it might be claimed that not all transhumanists face the metaphysical concerns articulated in this essay.

The problem with this response is that it is not clear that positions that reject the radical enhancements I have focussed on should count as transhumanist in the first place. We might argue as follows: moderate positions that repudiate uploading, and largescale replacement of the human person by artificial components leave the human person largely intact, albeit in an enhanced form. But the term 'transhumanist' strongly connotes the idea of using technology to advance *beyond* the category 'human'. So, on the face of it, it seems misleading to use this term for such moderate positions.

A better strategy is to argue that I have exaggerated the *degree* to which transhumanists support or promote radical technologies like uploading. It might be agreed, for example, that our epistemic situation would not warrant recommending the public to *undergo* radical enhancements. But it might be claimed that transhumanists are not committed to making this or similar recommendations. It might be claimed that they are only committed to the view that these technologies deserve investigation, should not be outlawed, and that they should be developed insofar as we can be confident they will benefit humans rather than harm them.

Two points can be made in reply. First if we allow moderate positions like this to count as 'transhumanist', there is once again a danger that the term will lose any interesting meaning. After all, there is nothing novel

about the idea that transhumanist technologies deserve investigation. Philosophy faculties and science fiction writers were engaged in this long before the transhumanist movement arose. Likewise, the view that these technologies should be developed *insofar as we can be confident they will benefit humans* can be agreed on by bioconservatives. So this defence of 'transhumanism' makes transhumanism much less interesting than it is usually taken to be.

Secondly, my strong impression—and I think it is a shared one—is that the enthusiasm of mainstream transhumanists for the radical enhancements discussed here far outstrips the moderate position just described. The time, effort, and funding that has been devoted to promoting these technologies implies that they deserve far greater investment than one would expect a proponent of the moderate position to accord them. If this impression is mistaken, then this seems to be due to a public relations failure on the part of the transhumanist movement.

There is one final response that that transhumanists might make that seems to me more promising. Transhumanists might concur that they are actively promoting the development and use of radical technologies like uploading; and they might accept that we cannot be confident that these technologies will have desirable, rather than having extremely undesirable effects. But they might argue that their position remains rationally defensible, because the only alternative, accepting human frailty and mortality, is itself so undesirable.

This is an important point. When contemplating the prospect of being uploaded or radically altered one can feel a degree of existential revulsion. Plausibly this is because we doubt that we will survive this procedure in a desirable state. But of course, we are faced with a similar problem in everyday life. Human existence proceeds under the constant threat of illness, calamity, and death. As such, transhumanists might concede that, given our weak epistemic situation, the technologies they promote *might* only hasten these or worse results, but argue that this risk is justified, since there is also a chance they will allow us to evade them.

This line of response depends on a very negative assessment of the natural life, and natural lifespan of human beings. It involves a change from the usual tone of transhumanism. For typically, transhumanism is presented in terms of the exciting, liberating prospects of technological advance. But this response suggests that it would be better characterised as an act of desperation in the face of death. Whether such an act is justified depends on wide-ranging issues that cannot be entered into here.

For those who are interested, Daniel Came's paper in Göcke and Meier-Hamidi (2018) is a good starting point.

Works Cited

Bayne, Tim, and David Chalmers. 'What is the Unity of Consciousness?' The Unity of Consciousness: Binding, Integration, Dissociation. Ed. Axel Cleermans. Oxford University Press, 2003. Print.

Bayne, Tim. The Unity of Consciousness. Oxford University Press, 2010. Print.

Carrol, Lewis. Alice's Adventures in Wonderland. London: Macmillan 1920 [1865]. Print.

Chalmers, David. The Conscious Mind: Towards a Fundamental Theory. Oxford University Press,1996. Print.

— The Character of Consciousness. Oxford University Press, 2010. Print.

Crick, Francis and Christof Koch. 'Towards a neurobiological theory of Consciousness.' Seminars in the Neurosciences. 2 (1990): 263-75. Print.

Dennet, Daniel, Consciousness Explained. London: Penguin, 1993 [1991]. Print.

Eccles, John. How the Self Controls its Brain. Berlin: Springer-Verlag, 1994. Print. Engel Gregory et al. 'Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems.' Nature 446 (2007): 782–6. Print.

García-Carpintero, Manuel. and Josep Macià. Eds. Two-Dimensional Semantics. Oxford University Press, 2006. Print.

Gauger, Eric et al. 'Sustained Quantum Coherence and Entanglement in the Avian Compass.' Physics Review Letters 106.4 2011. Print.

Göcke, Benedikt. 'Groundwork for a Dualism of Indistinction'. In After Physicalism. Ed. Benedikt Göcke. Notre Dame Press (2012): 263-294. Print.

Göcke, Benedikt and Frank Meier-Hamidi. Designobjekt Mensch: Die Agenda des Transhumanismus auf dem Prüfstand. Freiburg im Breisgau: Herder, 2018. Print.

Hameroff, Stuart and Roger Penrose. 'Consciousness in the universe: A review of the 'Orch OR' theory.' Physics of Life Reviews 11.1 (2014) 39-78. Print.

Hasker, William. 'Concerning the Unity of Consciousness.' Faith and Philosophy. 12.4 (1995): 532-47. Print.

Hobart, R. E. 'Free Will as Involving Determination and Inconceivable With-out It.' Mind 43.169, (1939) 1-27. Print.

Jackson, Frank. 'Epiphenomenal Qualia.' Philosophical Quarterly 32:127 (1982): 127-36. Print.

James, William. The Principles of Psychology. New York: Cosimo, 2007 [1890]. Print.

Kane, Robert. The Significance of Free Will. Oxford University Press, 1996. Print.

Kaplan, Mark. 'Decision Theory as Philosophy.' Philosophy of Science 50.4 (1983): 549-577. Print.

Kim, Jaegwon. Physicalism, or Something Near Enough. Princeton: Princeton University Press, 2005. Print.

Koons, Robert. and George Bealer. The Waning of Materialism. Oxford University Press, 2010. Print.

Kripke, Saul. Naming and Necessity. Harvard University Press, 1980. Print.

Kurzweil, R. The Age of Spiritual Machines: When computers exceed human intelligence. Viking Penguin, 1999. Print.

Laplace, Pierre Simon. A Philosophical Essay on Probabilities. Trans Frederick Wilson Truscott and Frederick Lincoln Emory. London: Chapman & Hall, 1902 [1814]. Print.

Lockwood, Michael. 'The Grain Problem.' Objections to Physicalism. Ed. Howard Robinson. Oxford University Press, 1993. Print.

McCulloch, W. and W. Pitts. 'A Logical Calculus of the Ideas Immanent in Nervous Activity'. Bulletin of Mathematical Biophysics, 7 (1943): 115–133.

Parfit, Derek. Reasons and Persons. Oxford University Press, 1986. Print.

Penfield, Wilder. Mystery of the Mind: A Critical Study of Consciousness and the Human Brain. Princeton University Press, 1975. Print.

Preist, Stephen. The Subject in Question: Sartre's Critique of Husserl in the Transcendence of the Ego. London: Routlege, 2000. Print.

Rumelhart, D., J. McClelland, and the PDP Research Group, Parallel Distributed Processing, vol. 1. Cambridge: MIT Press 1986. Print.
—Parallel Distributed Processing, vol. 2. Cambridge: MIT Press 1987. Print.

Sacks, Oliver. 'Neurology and the Soul.' The New York Review of Books, November, 1990. Print.

Schneider, Susan 'Why Property Dualists Must Reject Substance Physicalism.' Philosophical Studies 157 (2012): 61-76.

— 'Mindscan.' Science Fiction and Philosophy: From Time Travel to Superintelligence, Second Edition. Ed. Susan Schneider. Hoboken, NJ: John Wiley & Sons (2016) 260-76. Print.

Sherrington C. S. 'The Integrative Action of the Nervous System.' New Haven, CT: Yale University Press, 1947. Print.

Tye, Michael. Consciousness and Persons. Cambridge MA: MIT Press, 2003. Print.

Tononi, Giulio. 'Consciousness as Integrated Information: A Provisional Manifesto.' Biological Bulletin. 215.3 (2008): 216-242. Print.

Williams, D. C. 'On the Elements of Being I'. The Review of Metaphysics. 7.1 (1953): 3-18. Print.

Van Inwagen, Peter. An Essay on Free Will. Oxford: Clarendon Press, 1983. Print.

Zimmerman, Dean. 'From Property Dualism to Substance Dualism.' Aristotelian Society Supplementary Volume 84.1 (2010): 119-150.