Subjects, Experiences, and the Passage of Time

A neo-Parfittian account

Henry William Pollock

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds School of Philosophy, Religion, and History of Science

February 2020

The candidate confirms that the work submitted is their own and that appropriate credit has been given where reference has been made to the work of others.

Writing this thesis has been a joy, which I owe to several people. Firstly, Robin Le Poidevin and Helen Steward, my supervisors. They have been incredibly generous with their time and their knowledge, and I feel very lucky to have had their support. I owe much to Derek Parfit, whose writing had a transformative effect on my outlook, and whose words of encouragement spurred me on to pursue this PhD. I am also extremely grateful for the amazing opportunities given to me by my sponsors, the White Rose College of the Arts and Humanities. And lastly, thanks to my wonderful wife, Fran, without who I doubt this would have happened.

Abstract

The A and B-theorists of time disagree over whether time passes in reality. The B-theorist denies it does, and so, despite its successes, stands at an intuitive disadvantage. The A-theorist on the other hand is able to argue that our experience of time provides evidence of its passage. This 'argument from experience' expresses what I take to be the main motivation behind the A-theory. My aim is to provide the B-theorist with a response. The general thrust is that the argument from experience rests upon a mistaken view about the self – namely, Non-Reductionism. If we instead assume a Reductionist view, it should be rejected.

Derek Parfit argues that, given Reductionism, it can be an empty question whether persons persists through change. After defining and justifying Reductionism, I argue for a stronger claim: it is *always* an empty question whether persons – and their experiences – persist. That is to say, what we naturally describe as a single persisting person (or experience) could just as accurately be described as a series of distinct momentary persons (or experiences).

This claim is defended, then put to work against the argument from experience. Firstly, I argue it follows from this claim that we could not have veridical experiences of temporal passage. So, even if we do experience time as passing, we couldn't take this as evidence that time really does pass.

Finally, I propose a cognitive error account of temporal experience whereby although we *believe* we experience time as passing, this belief is false. I argue that the intuition to the contrary should be regarded as a side effect of a faulty, Non-Reductionist conceptual scheme. If we were to assimilate a Reductionist conceptual scheme instead, it would be impossible to conceptualise an experience of passage. In other words, time would not seem to pass.

Contents

…I seemed imprisoned in myself. My life seemed like a glass tunnel, through which I was moving faster every year, and at the end of which there was darkness. When I changed my view, the walls of my glass tunnel disappeared. I now live in the open air.

Parfit, *Reasons and Persons*

It is also worth investigating how time is related to the soul.

Aristotle, *Phyiscs*

For Bob

Introduction

Does time pass? To the uninitiated, the answer will seem obvious. But the question, in some form or another, is one that philosophers have been grappling with for millennia. In the early twentieth century, however, two papers were published that were to galvanise the debate: Albert Einstein's 'On the Electrodynamics of Moving Bodies' (1905), and John McTaggart's 'The Unreality of Time' (1908). McTaggart's paper teased apart two different ways of thinking about time. Within the 'A-series', events are ordered according to their changing relations to the present (pastness and futurity); within the 'B-series', events are ordered in terms of their unchanging relations to one another (earlier-than and later-than). Debate ensued over which of these conceptions more fundamentally describes reality. The A-theorist holds the common sense view that the passage of time and a changing present are objective features of reality.[1] The B-theorist on the other hand conceives of reality as a spatio-temporal 'block' in which 'now' is merely perspectival – analogous to 'here' in space – and from which the passage of time is absent. Einstein's paper meanwhile gave birth to relativistic physics, which, with its unification of space and time, and omission of an objective present, is widely thought to describe a B-theoretic block universe.

The B-theory is now the front runner, having accumulated a commanding array of supporting arguments. There does, however, remain a thorn in the B-theory's side: experience. Simply, time seems to pass. I take this to be the main motivation behind the A-theory, and is illustrated by the following oft-cited quotes:

> We must regard the feeling of 'becoming' as (in some respects at least) a true mental insight into the physical condition which determines it … Surely 'becoming' is a reality – or the nearest we can get to a description of reality. We are convinced that a dynamic character must be attributed to the external world. (Eddington 1928: 89, 94-95)

[1] How exactly the A-theorist unpacks this will depend on which version of the A-theory they espouse.

> I find it impossible to relinquish the sensation of a flowing time and a moving present moment. It is something so basic to my experience that I am repelled by the claim that it is only an illusion or misperception. (Davies 1995: 275)

> [Experience] is a defeater-defeater that overwhelms any B-theoretic arguments against the reality of tense. (Craig 2000: 138)

Temporal experience has in recent years become the focus of the debate between the A- and B-theorists. When A- and B-theorists disagree over whether time passes, what they are disagreeing about is the nature of change. For the B-theorist, change is mere temporal variation. For the A-theorist, an extra ingredient is need. A-theoretic change is essentially *dynamic*; it is the *replacement* of one state of affairs by a subsequent state of affairs.

The A-theorist's motivating intuition can be formulated as an inference to best explanation (see e.g. Baron *et al* 2015). Call this the *argument from experience*:

(A1) We have experiences as of A-theoretic change.[2]

(A2) The best explanation for (A1) is that A-theoretic change occurs in reality.

(A3) Therefore, A-theoretic change occurs in reality.

My aim is to provide the B-theorist with a response to this argument. It is my contention that the ostensible plausibility of this argument results from an erroneous conception of the self and its identity over time.

Few writers have been as influential in the field of personal identity as Derek Parfit. Parfit's Reductionism is well known not only for being revisionist in its claims about the self, but also for its far-reaching implications regarding rational decision-making and ethics. I want to show that it can also weigh in on the debate between the A- and B-theorists of time. The idea is that a properly Reductionist account of the self informs our understanding of temporal experience; and that this in turn undermines the A-theorist's argument from experience.

There are five chapters. Chapters 1 to 3 develop and defend what I am calling a 'Neo-Parfittian' metaphysics of personal and experiential identity. Chapters 4 and 5 then put

[2] 'As of' rather than 'of' so as not to presuppose the experience to be veridical.

this metaphysics to use against either premise of the A-theorist's argument from experience.

The purpose of Chapter 1 is to define Reductionism as propounded by Derek Parfit (1987), and demonstrate Non-Reductionism to be untenable. Reductionism is a family of claims about what personal identity consists in (for example, psychological continuity). However, despite its influence, a satisfactory definition of Reductionism remains elusive. This is no doubt due in large part to the fact that Parfit's own offerings are somewhat lacking in precision. I begin by taxonomising the various possible positions. The shortcomings in Parfit's definitions are then highlighted by comparing them to those of another closely-related family of views: Criterialism (and Anti-Criterialism).

Parfit gives us two paradigm examples of Non-Reductionism: the Cartesian View and the Further Fact View. Arguments against these positions are considered, and, where needed, shored up. It is found that these arguments rule out all but one of the positions in the taxonomy: the view that personal identity consists in empirical relations holding between empirical entities. This is settled upon as our definition of Reductionism; all others positions in the taxonomy are Non-Reductionist.

In Chapter 2 I argue that, given the truth of Reductionism, it is an empty question whether persons, and their experiences, persist. A question is empty when disagreements over how to answer it are merely verbal. Parfit employs his Combined Spectrum thought experiment to argue that a person's diachronic identity can be indeterminate, and that when this is the case, it is because the question of that person's identity is empty.[3] This is an intriguing claim, since indeterminacy in thought experiments like the Combined Spectrum is generally held to arise as a result of vagueness; but vagueness is conspicuous in its absence from Parfit's discussion.

A way forward is found in Ted Sider's (2011) treatment of verbal disputes, which allows us to accommodate vagueness in the Combined Spectrum while retaining the conclusion that identity questions can be empty. Having done so, I then argue that a stronger conclusion is warranted: it is *always* an empty question whether a person persists through change. I extend this claim further still, arguing that, on the assumption that experiences

[3] An example of such a question of identity would be: 'does person p1 persist (retain their identity) through changes [x, y, z]?'. All discussion of identity will be of *diachronic* identity, unless otherwise stated.

are individuated by the persons *qua* subjects who have them, the same can be said of experiences: although it is natural to describe a continuous experiential episode as a single unified experience, it would be just as metaphysically accurate to describe it as a series of distinct momentary experiences. A potential reply to this 'Neo-Parfittian' claim in the form of Barry Dainton's (2008) phenomenal continuity account of personal identity is considered, but is found to be a version of Non-Reductionism and disregarded accordingly.

Chapter 3 defends the Neo-Parfittian claim against the arguments of Matthew Soteriou (2013) and Ian Phillips (2011a). Both argue for a 'holistic' conception of experience according to which 'whole' extended experiences are metaphysically prior to their momentary parts. If true, it would mean that we can only make sense of momentary experiences as *part* of extended wholes, and the Neo-Parfittian claim would be false: we *couldn't* just as well describe an extended experience as a series of distinct momentary experiences. Both writers argue that motion perception presents certain puzzles that can only be solved from within a holistic framework. I show that their arguments are unsuccessful, and that it is in fact the view they attack which is placed to make sense of the phenomena in question.

I first examine Soteriou's argument. I show that the puzzle Soteriou raises can only *be* a puzzle if we have already presupposed the 'atomistic' conception he opposes to be false. If instead we assume it to be true, there is no mystery whatsoever. Phillips's argument is examined next, and found to fare no better. The ostensible paradox he raises is indeed troublesome. However, I show that his solution has unacceptable neurological consequences, and runs together two importantly distinct neural mechanisms. I present a much simpler solution to the paradox which avoids these problems, and does so without the need for the holistic conception of experience. We need be worried by neither puzzle.

In Chapter 4, the Neo-Parfittian claim is put to use against (A2) of the A-theorist's argument from experience, which claims that the best explanation for our experiences as of passage is that time passes in reality. I counter by arguing that, given the Neo-Parfittian claim, we would not be justified in believing that we have *veridical* experiences of passage; and so *even if* time did pass, we couldn't take this to explain the experience claimed by (A1).

Before outlining the counter-argument, a word on terminology. My use of 'event' and 'state' throughout the thesis will be somewhat idiosyncratic. Usually, when writers use the term 'event' they mean something that is processual, or something that changes over time; by 'state', writers usually mean something that is unchanging. Normally, then, states can be said to have temporal extension. For example, one could be said to be in a *state* of happiness over some period. When I say 'event', however, I merely mean something with temporal extension. So if I feel happy for some period, this would, according to my terminology, be a happiness *event.* When I say 'state' I will be talking about something without temporal extension. A state can then be thought of as, if one accepts such talk, a 'time-slice' *of* an event (for this reason I may sometimes talk of states *of* events). So a happiness event will comprise happiness states that obtain *during*, or *in* the event (I accept both ways of talking). A state is always *at* a time, whereas an event is always *over* time.[4]

Now for the counter-argument of Chapter 4. The outline is as follows. I argued in Chapter 2 that it is an empty question whether experiences have temporal extension. If it is an empty question whether experiences have temporal extension, any theory that needs to posit extended experiences to account for our perception of extended events must be rejected. We must be able to account for our perceptual awareness of extended events within momentary experiential states. If our awareness of extended events must be accounted for within momentary experiential states, our awareness of extended events must be *derived* from our awareness of instantaneous states obtaining *during* events. Given that our perception of events is derivative in this way, A-theoretic change (A-change) cannot be an object of experience; even if change has an A-theoretic structure, we could not experience it as such. So, even if we allow that A-change is represented in experience, we must accept that it is non-veridical. (A2) is refuted.

Finally, I reply to (A1) in chapter 5. I propose a cognitive error account of temporal experience whereby although we *believe* we have experiences as of passage, this belief is false. I begin by softening intuitions against the idea that we could make mistakes about the content of our own experience. Some people think it is impossible to be mistaken in this way: when it comes to one's own experience, what *seems* is what *is*. I begin by showing that this cannot be right as a general principle since there are times when we *do* make such mistakes. I then consider a version of this principle restricted to temporal

[4] A persisting object could therefore be thought of as a spatio-temporal event; a photograph of that object would capture an image of a *state*.

experience. By framing it within the debate between conceptualists and nonconceptualists, however, I show that a proponent of this restricted principle would face a dilemma, finding enemies on both sides of the debate.

The B-theorist may therefore reject the A-theorist's motivating intuition that experience represents time passing. Provided, that it, they can explain how it is we come to make such a mistake. I propose that this mistake results from our Non-Reductionist conception of the self. We naturally suppose that our persistence is metaphysically significant; that questions of identity are not empty. This conception of the self is built into our conceptual scheme in the form of first person pronouns. Because Reductionism is true, however, a conceptual scheme which lacked the concept of diachronic personal identity would be no less metaphysically accurate than ours. We could use 'we' to refer to our temporally extended selves rather than 'I'. The intuition that I experience A-change necessarily presupposes that I am the same subject that experiences both states of affairs. Without the concept of personal identity (and experiential identity), it would make no sense to suppose that I could experience A-change. Such an experience would be impossible to conceptualise. And so it would be impossible to conceptualise an experience *as of* A-change. We would be faced with a certain phenomenology, but we could not characterise this phenomenology as time passing. (A1) is refuted.

My overall conclusion is that neither (A1) nor (A2) should be accepted, and, as such, that the A-theorist's argument from experience fails. We must resist our A-theoretic intuitions; experience gives us no reason to believe that time passes.

Chapter 1: Personal Reductionism

The aim of this chapter is to define Reductionism as propounded by Derek Parfit, and to show that Non-Reductionism should be rejected. Few writers have been as influential in the field of personal identity as Derek Parfit. Yet despite this, a satisfactory definition of Reductionism remains elusive. Parfit makes several claims that fall under the banner of Reductionism: that identity does not matter in survival; that identity can be indeterminate; and so on. I focus on his ontological claim: the claim about what personal identity *consists in.*[1]

When considering the question of what personal identity consists in, there are many different ways one might respond. I begin by taxonomising the various positions, before showing why Parfit's own definitions are unsatisfactory. I then critically evaluate the arguments populating the literature that would count against certain positions in the taxonomy. It is shown that not only does Parfit fail to offer a watertight case against Non-Reductionism, but that there are other views that should count as Non-Reductionist that he fails to identify. A superior definition of Reductionism is settled upon, and the case against Non-Reductionism is shored up. It is concluded that Reductionism is the only tenable position.

1.1 Taxonomising the positions

Broadly speaking, we can divide the different responses to the question of what identity consists in into two sorts, which I shall call *informativist* and *primitivist*. An informativist thinks something informative (non-circular, non-trivial) can be said about what grounds personal identity. Their explanans (what identity consists in) contains concepts that aren't included in, or equivalent to the concepts in, the explanandum ('personal identity').[2] The primitivist denies this. For the primitivist, the explanans contains only those concepts that

[1] Following Parfit (1999), I will refer to the ontological claim simply as 'Reductionism'.
[2] Whenever I write 'identity' I am referring to the diachronic relation, unless otherwise stated.

are included in, or equivalent to, those in the explanandum. Their response is uninformative.

There are two main ways an answer can be informative. One sort of answer tells us something informative about the *entity* that's involved. For example, one might say that personal identity consists in the persistence of an empirical entity, such as a body.[3] Similarly, an informativist might say that identity consists in the persistence of something non-empirical, like a separately existing immaterial soul-like entity, or Cartesian Ego. When a position is informative about the entity involved, we can call it 'entity informativism', or 'E-informativism' for short. We can then differentiate between the empirical and non-empirical versions as 'E-informativism$^{\mathbf{E}}$' and 'E-informativism$^{\mathbf{NE}}$' respectively.

The other sort of informativist answer says something informative about the relation itself. We may call this position 'relation informativism', or 'R-informativism' for short. Again there are empirical and non-empirical variants. The typical empirical variants are physical continuity and/or connectedness (C/C), and psychological C/C. It might help to say a little about these relations. Firstly, 'continuity' is transitive, whereas 'connectedness' is not: even if person p1 at time t1 is not *connected* to p3 at t3, p1 at t1 will be *continuous* with p3 at t3 provided that they are connected to p2 at t2, who is in turn connected to p3 at t3. I (now) can be physically (or psychologically) *continuous* with an earlier self even in the absence of any direct physical (or psychological) *connections,* provided there is a chain of physically (or psychologically) connected selves between us. In this way, continuity can be used to explain how, for example, I (now) can be considered the same person as four-year-old-Henry, despite the fact that my body is now made up of new cells, and my psychology is radically different.

There still remains the question of what exactly *physical* and *psychological* C/C *are*. We should be careful not to impose too strict a definition here – after all, we are trying to chart the *possible* views on personal identity, and there may well be different views about what counts as a physical or psychological connection (and therefore what counts as an instance of continuity). Parfit's writing (1987: 203-204) suggests that he thinks physical connectedness is simply a matter of being made up of the same physical stuff, such that I am physically connected to a past self in virtue of having the same cells etc. However,

[3] I am here using 'empirical' in the sense of discoverable through experience.

this is to define physical connectedness in terms of *identity*. An account of personal identity along these lines therefore fails to tell us anything informative about the *relation* that grounds personal identity. It only tells us about the *entities* involved – that they are made up of physical stuff. It merely says that *personal* identity consists in *bodily* identity. This family of views – those that are informative about the entity but uninformative about the relation – will be addressed shortly.

For an account based on physical C/C to be informative about the *relation*, it cannot define physical connectedness in terms of bodily persistence.[4] It cannot employ the notion of *quantitative,* or *numerical* identity – that one and the same body (or body part) is existing at two different times. What it can do, however, is employ the notion *qualitative* sameness, or exact similarity. On this view, I am physically connected to a past self in virtue of having certain distinctive physical properties in common with that person (properties that don't presuppose numerical identity). If we could give a complete physical description of my body at t1, much of this description would be rather generic (two arms, two legs etc.). But at a greater level of detail it would describe many features that were peculiar to me, even down to the arrangement of cells. If we were to give a second such description a few moments later, we would find there to be many similarities between the two.[5] With many of the descriptive peculiarities, there would be an exact similarity. It would then be in virtue of these similarities that physical connectedness holds.[6]

This notion of connectedness can then be transposed to *psychological* connectedness. The classic example of a psychological connection is memory: if p2 at t2 remembers the experiences of p1 at t1, then p1 and p2 are psychologically connected. But other psychological connections could be beliefs, hopes, desires, and so on. In short, we may

[4] I assume that 'persistence', 'continued existence', and other such concepts presuppose identity.
[5] Connectedness holds as a matter of degrees: the more similarities, and the more distinctive those properties are, the higher the degree to which the two persons would be connected.
[6] Additional conditions could be added. For instance, one might want to say that p1 and p2 must be spatiotemporally contiguous in order to be physically connected. Such a view would then rule out true teleportation. Parfit seems to assume such a condition when he says '[t]hose who believe in the Physical Criterion would reject Teletransportation' (1987: 204). Although this would be true on many plausible interpretations of the physical criterion, other interpretations need not be so strict. I will assume throughout however that these views include a non-branching clause: whatever relation is taken to ground identity, that relation cannot take a branching form, since a single person at t1 cannot be identical to two distinct persons at t2.

say that psychological connectedness holds between p1 and p2 when there are certain distinctive similarities between the psychologies of p1 and p2.

So much for the empirical variants of R-informativism. The non-empirical variant claims that personal identity consists in immaterial C/C – that is, continuity and/or connectedness of the spooky stuff that (one might think) Cartesian Egos are made of. As Richard Swinburne writes:

> We may say that there is a stuff of another kind, immaterial stuff, and that persons are made of both normal bodily matter and of this immaterial stuff but that it is the continuity of the latter which provides that continuity of stuff which is necessary for the identity of persons over time. (Swinburne 1984: 27)

We can now differentiate between the empirical and non-empirical versions of R-informativism as 'R-informativism^{E}' and 'R-informativism^{NE}' respectively. These four different types of informativist position can be plotted onto a table, which can be populated with examples where applicable.

		Relation	
		Empirically informative	**Non-empirically informative**
Entity	**Empirically informative**	Physical C/C of bodies Psychological C/C of brains	n/a
	Non-empirically informative	Psychological C/C of Cartesian Egos	Immaterial C/C of Cartesian Egos

Figure 1: taxonomy i

It will be useful to make some clarifying remarks about the examples given in Figure 1. Take 'physical C/C of bodies' for starters. Physical C/C holds between physical states, and in this instance those physical states are the physical states of bodies. The example is empirically informative about both the relation and the entity. So the example fits straightforwardly in the E-informative^{E}/R-informative^{E} box (top left). But psychological C/C is not so straightforward. Psychological C/C holds between psychological states. But there is disagreement over what sort of entity psychological states are attributable *to*. A

substance dualist will think that psychological states are attributable to immaterial mental entities like Cartesian Egos, whereas a materialist will think that brains (or bodies, or organisms etc.) are the bearers of psychological states. So psychological C/C might be attributed to empirical or non-empirical entities. For this reason, psychological C/C features in both boxes in the left hand column (E-informative$^{\mathbf{E}}$/R-informative$^{\mathbf{E}}$ and E-informative$^{\mathbf{NE}}$/R-informative$^{\mathbf{E}}$).

I have left the E-informative$^{\mathbf{E}}$/R-informative$^{\mathbf{NE}}$ (top right) box as n/a; no view can be empirically informative about the entity but non-empirically informative about the relation. It might be objected that there *is* a type of view that belongs here. The view I have in mind might be called the 'divine bundle theory'. According to this view, a person is just a bundle of physical and/or psychological states, and these states are unified over time as a matter of divine will, or suchlike. Bundles of empirical states are connected merely because God decided so. The thought goes that we would be saying something informative about the relation – *viz.* that it is a divine sort of connectedness – but we are also admitting that it is beyond our empirical grasp. However, although there is a sense in which this view says something informative about the relation, it is not informative in the right sort of way. It is informative only insofar as it says something about the *cause* of the relation: God causes the unification relation to hold. But it doesn't tell us anything about the relation itself. After all, someone might hold that personal identity consists in physical continuity between bodies, but nevertheless hold that this physical continuity holds as a matter of divine will. Or someone might hold that persistence consists in the divinely willed immaterial continuity between Egos. These two views respectively say something empirically and non-empirically informative about the relation involved in persistence. Hence, knowing that the relation that grounds persistence holds as a matter of divine will is, in itself, insufficient to say whether or not a view is empirically or non-empirically informative about that relation. We are now in a position to see why the divine unification theory fails to count as a R-informative view. It says something informative about *why* the relation holds, but not about what the relation itself actually is.

There is a *prima facie* case for supposing that entity informativism and relation informativism go hand-in-hand. If you think persons persist by virtue of the persistence of their physical bodies (E-informativism$^{\mathbf{E}}$), then it is natural to also think that physical C/C is the grounding relation (R-informativism$^{\mathbf{E}}$). Or if you think that persons persist by virtue of the persistence of an immaterial substance (E-informativism$^{\mathbf{NE}}$), then it would

make sense to hold that immaterial C/C is the grounding relation (R-informativism^{NE}). So it is plausible to suppose that someone who claims something informative about the entity will do the same regarding the relation.

And indeed they might do. But there is nothing *forcing* them to. One might think that something informative can be said about the entity *only*. For example, someone might think that a person persists by virtue of the persistence of a Cartesian Ego, any yet contend that the persistence of Egos is brute. Similarly, someone might think that we persist by virtue of the persistence of our bodies, but hold that bodily persistence is brute. Such positions are uninformative about the grounding relation, but informative about the entities it holds between. We can therefore add a column to our table for those positions that say something informative about the entity but not the relation itself. Following suit, we can call these positions 'R-uninformative'.

		Relation		
		Empirically informative	**Non-empirically informative**	**Uninformative**
Entity	**Empirically informative**	Physical C/C of bodies Psychological C/C of brains	n/a	Brute bodily persistence
	Non-empirically informative	Psychological C/C of Cartesian Egos	Immaterial C/C of Cartesian Egos	Brute Cartesian Ego persistence

Figure 2: taxonomy ii

It does seem, on the other hand, that if we are to say something informative about the relation we must say something informative about the entity. This is because if we say something informative about the relation we will be attributing to the entity the states between which the stipulated relation holds. For example, if you think that persistence consists in physical continuity, you will have to accept that the grounding entity – the entity involved in whatever grounds identity – has physical states. And that is to say something informative about the entity. Admittedly, a *more* informative answer could be given. Knowing the sort of states the entity has doesn't tell us everything there is to know

about the entity. But still, in saying something informative about the relation, something informative about the entity will follow.

It might not seem obvious that this must be the case. Some relations, one might think, do not entail anything informative about the entity. For instance, it might seem plausible for someone to hold that causality is the grounding relation whilst having nothing informative to say about the entity. But although they might not *think* that they have anything informative to say about the entity, they must at least admit that, whatever the entity is, it has causal states (i.e. states that figure in causal relations). Now, if *all* candidate entities have causal states, this wouldn't tell us anything informative about the entity. But there are surely potential candidate entities that don't have causal states. Someone might, for example, claim that the spooky states of Cartesian Egos are causally unrelated. This would be an odd view, for sure. But if someone thinks that personal identity consists in brute Cartesian Ego persistence, they needn't think that *any* relation holds between the states of Egos except *identity*. As such, if someone claims that the entity that grounds personal persistence has causal states, they will be claiming something informative about that entity.

That leaves one last combination that hasn't yet been discussed: the position whereby nothing informative can be said about either the entity or the relation. This amounts to a full-blooded primitivism, which Parfit calls the Further Fact View (1987: 210). According to the primitivist, the fact of a person's persistence does not consist in any other facts at all. It is a 'further fact'.[7]

It should be noted that a primitivist might be able to claim that personal identity consists in the persistence of an ego. This might seem like an inconsistent position. Primitivists, after all, deny that identity consists in any other facts. But they *could* claim that identity consists in the persistence of an ego, *provided they weren't using 'ego' in the same way as the E-informativist*^{NE}. Rather than holding that egos are immaterial entities, they would have to be using 'ego' in such a way that made their claim only trivially true. They might for instance use 'ego' merely as a placeholder for 'person', or to mean something like

[7] There is a *sense* in which the Cartesian Views posit a 'further fact'. Once we have all the physical and psychological facts to hand, the Cartesian will still say that there is a 'further' fact to a person's identity – i.e. a non-empirical fact about an immaterial entity. However, the Cartesian nevertheless claims that personal identity *consists* in this fact. And so it is not a further fact in the sense intended by Parfit. The primitivist denies that personal identity consists in any other facts. This is why the primitivist's view is called the *Further Fact View*.

'essence of a person' – which is then defined circularly in terms of personal identity. We can now introduce the 'E-uninformativism' row on our table, and populate it accordingly:

		Relation		
		Empirically informative	**Non-empirically informative**	**Uninformative**
Entity	**Empirically informative**	Physical C/C of bodies Psychological C/C of brains	n/a	Brute bodily persistence
	Non-empirically informative	Psychological C/C of Cartesian Egos	Immaterial C/C of Cartesian Egos	Brute Cartesian Ego persistence
	Uninformative	n/a	n/a	Primitivism

Figure 3: taxonomy iii

1.2 Reductionism and Criterialism

Now that we can see the full range of positions, we can take a look at how Parfit defines Reductionism and Non-Reductionism. As we will see, Parfit's writing sometimes suggests a way of defining the terms that would bring them in line with another family of views: Criterialism (and Non-Criterialism). However, when we do so we get the wrong results: positions Parfit explicitly characterises as Non-Reductionist are in fact Criterialist positions. This section will expose the flaws in Parfit's definitions, and take some initial steps towards finding satisfactory replacements.

Parfit (1987: 210) defines Reductionism as the view that:

> R1: 'the fact of a person's identity over time just consists in the holding of certain more particular facts.'[8]

[8] Parfit adds an optional second condition for Reductionists: that the facts in which identity consist must be describable without presupposing the existence of persons. There are further problems with this second condition, and, since Parfit allows that Reductionists need not commit themselves to it, I shall ignore it. For discussion, see Hummel (2017).

A view is then defined as Non-Reductionist if it rejects R1 (*ibid.*).

There has been much debate over whether Parfit's use of the term 'reductionism' is warranted (e.g. Cassam 1989; Garrett 1991; McDowell 1997; Shoemaker 1997; Blackburn 1997; Parfit 1999; Behrendt 2003).[9] It is now common to find philosophers using the terms 'Criterialism' and 'Anti-Criterialism' instead, which *prima facie* appear to be doing the same work as Parfit's 'Reductionism' and 'Non-Reductionism' respectively. Duncan writes:

> Some philosophers think that personal identity through time *consists in* something, [and] that there are informative necessary and sufficient conditions – i.e. criteria – for personal identity through time. These philosophers are *criterialists*. (Duncan 2014: 283)

Similarly, Olson writes:

> I propose that a proper criterion is one that completes this formula:
>
> Necessarily, if *x* is a human person at *t* and *y* exists at t*, *x* = *y iff* and because…
>
> where the word "because" expresses the logical or metaphysical dependence […] Criterialism, then, is the claim that some such criterion is true. (Olson 2012: 49)

According to Criterialists, then,

C1: Personal identity consists in something (other than itself), and

C2: There are non-trivial necessary and sufficient conditions for personal identity.

I take C1 and R1 to be equivalent for our purposes. Although Parfit does not stipulate an equivalent of C2 in his definition, it would appear that he implicitly assumes it. We can see this assumption at work is his characterisation of a Reductionist psychological criterion:

[9] Many writers think that 'Reductionism' is something of a misnomer. The aim of this chapter is not to defend the use of 'Reductionism' as a technical term. Rather, my aim is to find a satisfactory definition and defence of the position Parfit *calls* 'Reductionism' – whether or not that name is warranted.

> X today is one and the same person as Y at some past time *if and only if* (2) X is psychologically continuous with Y, (3) this continuity has the right kind of cause, and (4) it has not taken a 'branching' form. (5) Personal identity over time just consists in the holding of facts like (2) to (4). (Parfit 1987: 207, emphasis added)

The 'if and only if' sets out the necessary and sufficient conditions as required by C2. At first blush, then, it seems that there is no real difference between Reductionism and Criterialism. The immediate problem with this diagnosis is that we know from Parfit's examples of Non-Reductionism that some Non-Reductionists should be classified as Criterialists. Parfit tells us that there are two sorts of Non-Reductionist view: the Further Fact View and the Cartesian View. The Further Fact View says that a person's identity consists in no other facts – i.e. that it is a primitive fact. This rejects R1, and so, as we should expect, counts as a version of Non-Reductionism. With the Cartesian View, however, we get the wrong result. Someone who holds the Cartesian View thinks that a person's identity is tied to that of an Ego, and so will hold one of the entity informativist^{NE} views. But none of these views reject R1; they all say that identity consists in 'certain more particular facts' – fact about Egos. Indeed, looking at Figure 3 we can see that *every* view except for E-uninformativism/R-uninformativism (*i.e.* the primitivist's Further Fact View) claims that identity consists in other more particular facts. So by R1, the Cartesian View is classified as Reductionist. This is not what Parfit wants.

Things get more muddled when we look at Parfit's attempts at clarification. He later says that Reductionism is the view that:

> R2: 'Our identity over time consists in physical and/or psychological continuity.' (Parfit 1999: 217)

Now that we have taxonomised the positions, we can see that it is possible to say something non-empirically informative about the entities, whilst saying something empirically informative about the relation that holds between them: it is possible for a Cartesian to hold that identity consists in psychological continuity. According to R2, this would mean they are a Reductionist. Again, we get the wrong result. What's causing this problem is that R2 defines a view as Reductionist in terms of the *relation* it posits, whereas the Cartesian view is defined in terms of the *entities* it posits. It is clear now that satisfactory definitions of Reductionism and Non-Reductionism need to account for both.

If we assume that the two Non-Reductionist views Parfit discusses (*viz.* the Cartesian View and the Further Fact View) are exhaustive of Non-Reductionism, we end up with a definition of Non-Reductionism as the view that identity consists in facts involving Egos, or else in no other facts at all; and Reductionism as the view that identity consists in other facts not involving Egos (Pollock 2018). Having distinguished between facts about entities and facts about relations, we can see that this would classify two positions as Reductionist: the E-informativist$^{\mathbf{E}}$/R-informativist$^{\mathbf{E}}$ view (top left box), and the E-informativist$^{\mathbf{E}}$/R-uninformativist view (top right).

The rest of this chapter will critically examine arguments in the literature that are effective against the various positions in our taxonomy. It will be shown that we should not take the Further Fact View and the Cartesian View to be exhaustive of the Non-Reductionist positions, as the considerations that count against the primitivist also count against the E-informativist$^{\mathbf{E}}$/R-uninformativist. In other words, any view that says the fact of a person's identity consists in the *brute persistence of any kind of entity* should be classed as Non-Reductionist, and disregarded as such. A satisfactory definition of Reductionism and Non-Reductionism will finally be settled upon, with Reductionism being the only tenable position. We begin by looking at the objections against the Cartesian and Further Fact Views.

1.3 Arguments against the Cartesian View

We can begin with the tried-and-tested arguments against the Cartesian View. In general terms, the case against the Cartesian View is epistemological: we have no evidence for Cartesian Egos, so we shouldn't believe they exist. There are two versions of this argument, which apply to different versions of the Cartesian View.

One version of the Cartesian View is what Parfit calls the 'Featureless Cartesian View'. Proponents of the Featureless Cartesian View hold that Egos are devoid of empirical features. Someone who holds this view would either be R-uninformativist or R-informativist$^{\mathbf{NE}}$. That is to say, they would hold that personal identity consists in the brute persistence of Egos, or else that personal identity consists in the immaterial continuity of Egos. Parfit's argument against these views follows the objections of Locke (1694). Since featureless Egos are devoid of empirical features (e.g. psychological states), it would be *impossible* to have any evidence of their existence or persistence. Although we are aware of psychological continuity, we cannot attribute this to an Ego if we are supposing that

Egos lack empirical features. Because of this, it would be impossible to know if identity ever held between two persons. Egos could swap between bodies and minds; or they could exist only momentarily, with one being replaced by another in a continual cycle. So what we intuitively reidentify as a single persisting person could actually be a series of distinct persons. An account of personal identity that makes it impossible to know whether or not identity ever holds between two persons should be rejected.

Parfit does accept, however, that there *could* have been evidence in favour of the Cartesian View. Or rather a *version of* the Cartesian View: the R-informativist$^{\mathbf{E}}$ version. According to the R-informativist$^{\mathbf{E}}$ view, empirical relations – i.e. psychological C/C – hold between Egos, since Egos are the bearers of psychological states. Parfit imagines a Japanese woman who claims to remember being an ancient Celtic Warrior. Although we would be rightfully sceptical of her claim, Parfit imagines that she is able to make certain predictions though would otherwise have been impossible to make. For example, she is able to say where certain settlements were, and successfully direct archaeological digs to places where she remembers hiding particular artefacts that she can accurately describe, and so on. If we lived in a world where this was commonplace, it might be reasonable to conclude that the carrier of memory was not the brain but some spooky substance or entity that could pass from one brain to another over the generations. In such a world, there would be empirical support for the R-informativist$^{\mathbf{E}}$ version of the Cartesian View. But we don't live in a world like that. In our world there is no such evidence. And so, just as we have no reason to believe that, say, fairies exist, we have no reason to believe that Cartesian Egos exist. The R-informativist$^{\mathbf{E}}$ version of the Cartesian View should therefore also be rejected.

Between these two arguments, all versions of the Cartesian View are ruled out. These arguments are well discussed, and I will assume their efficacy. The E-informativist$^{\mathbf{NE}}$ (middle row) positions in Figure 3 may therefore be discounted.

1.4 Arguments against the Further Fact View

1.4.1 The Shoemaker-Parfit argument

David Shoemaker (2002) offers an objection against the primitivist's Further Fact View which builds on Parfit's Combined Spectrum thought experiment. Parfit asks us to imagine a range of operations in which the distinctive physical and psychological (*viz.*

empirical) properties of person p1 (who, in this case, is Derek Parfit) are replaced with those of p2 (Greta Garbo). The idea is to test our intuitions about how much change a person can go through before they can no longer be considered to be the *same* person. At the near end of the spectrum, some minutiae of p1's empirical properties – say, a cell and a vague aspect of a distant memory – are replaced with those of p2. In the next case along, a trivially larger amount of p1's empirical properties are replaced with those of p2. This continues until we reach the far end of the spectrum, where we have an operation in which *all* of p1's distinctive empirical properties are replaced with those of p2.

It is important to note that the spectrum we are imagining is not a *series* of operations. We are not to imagine a single experiment consisting of multiple operations conducted one after another, each changing some further trivial aspect of p1's empirical properties. The spectrum does not represent a single experiment in which p1's empirical properties are progressively replaced. Rather, the operations are, as it were, *parallel* experiments, in the sense that each case on the spectrum represents a distinct possible experiment in which only one operation is made. At the near end of the spectrum we are to imagine some trivial aspect of p1's empirical properties are replaced in a single operation; at the far end we are to imagine all of p1's distinctive properties are replaced in a single operation.

The Combined Spectrum will be discussed in greater detail in Chapter 2. For now, though, it will suffice to say that Parfit's main target is the intuition that identity is always determinate. If Parfit's argument is successful, it would show that there are conceivable cases where we can't give a yes or no answer to the question of whether a person persists. Shoemaker takes this conclusion and uses it to construct an argument against the primitivist's Further Fact View, which we may call the Shoemaker-Parfit argument.

(SP1) If the Further Fact View is true, identity must always be determinate.

(SP2) The Combined Spectrum thought experiment shows that identity can be indeterminate.

(SP3) Therefore, the Further Fact View cannot be true.

The argument is unsuccessful, however. Neither premise of this argument need be accepted by the primitivist.

In support of (SP1), Shoemaker says of the primitivist's 'further fact':

> [T]he further fact either obtains or it does not. That is, questions of identity always have determinate answers: either "I" exist of "I" do not. The reason for this entailment is that, according to these views, our identity involves *one more thing* that the more particular (numerous, perhaps) facts about brains, bodies, etc. viz. it involves *a* fact about *the* entity which is me, some one thing that is distinct from the mere agglomeration of my material constitution. So given this further fact is *a fact*, it either obtains or it doesn't. (Shoemaker 2002: 152)

Although the primitivist *may* accept this story, I see no reason why they should be forced to. Yes, the primitivist believes the fact of a person's identity is a further fact. That is just to say they think that, when it is a fact that a person persists, this fact does not consist in other facts. Similarly, the primitivist thinks that, when a person does *not* persist, this fact consists in no other facts. Again, this fact would be a 'further fact' – a further fact about identity determinately *not* holding.

Now, if this is what the primitivist can say about instances where identity either does or doesn't hold, why can't they tell the same sort of story about instances where identity neither does hold nor doesn't hold? If they can say that there are further facts about determinate identity, why can't they say there are further facts about indeterminate identity? There doesn't seem to be anything preventing a primitivist from saying: there can be cases where it is neither true nor false that a person persists; that when this is the case it is a *fact* that identity neither does nor does not hold; and that this fact does not consist in other facts – that it is a 'further fact'. Indeed, if someone *were* to hold such a view, what does Shoemaker think we should say their position is? The position is undeniably primitivist, since it say that the fact of indeterminate identity consists in no other facts. So presumably Shoemaker would have to say that it is the wrong *sort* of primitivism – that the primitivist we have been discussing *only* thinks that the further facts of identity are *determinate* facts. But that would be to impose an unwarranted restriction on the definition. The point is that someone who thinks facts about determinate identity consist in no other facts (i.e. the primitivist as defined by Shoemaker) *can also* contend that facts about *indeterminate* identity consist in no other facts. The truth of the Further Fact View would not therefore entail that identity must always be determinate. (SP1) is false.

The primitivist doesn't *need* to accept that identity can be indeterminate in order to refute Shoemaker's argument, though. They might instead focus their attention on (SP2) – that the Combined Spectrum shows that identity can be indeterminate. A primitivist who rejects indeterminate identity will contend that there is a sharp cutoff somewhere along the spectrum between identity and non-identity; that there is a borderline between a pair of adjacent cases where on one side we have a clear case in which the original person persists through the change, and in the next we have a clear case in which they are replaced by a distinct new person.

Parfit objects that this position is indefensible. Only someone who believes in the existence of separately existing Cartesian Egos (i.e. a non-empirical informativist about the entities that ground personal identity) is entitled to believe that there is a sharp cutoff on the spectrum. As the primitivist doesn't believe that such entities exist, they cannot hold that a sharp cutoff exists:

> There are some people who believe that our identity must be determinate, though they do not believe that we are separately existing entities distinct from our brains and bodies, and our experiences. This view I believe to be indefensible. What explains the alleged fact that personal identity is always determinate? […] if we are not separately existing entities, how could there be such a borderline? What could make it true that, in one case, the resulting person would be me, and in the next he would not be me? What would the difference consist in? (Parfit 1987: 239)

The thought is that if Cartesian Egos existed (and, presumably, weren't ontically vague entities), there would be a fact of the matter as to when the original Cartesian Ego ceased to exist and a new one began.[10] And this fact would make it true that, in one case on the spectrum we could have a clear case of identity, and in the next case a clear case of non-identity. But for someone who doesn't believe that persons are separately existing entities, nothing could make it true that there is a sharp cutoff. As Parfit asks, what could the difference between identity and non-identity consist in?

[10] Proponents of ontic vagueness say that the world itself can be vague, and that vagueness is not (only) due to our use of language etc. If a Cartesian thought that Egos were ontically vague entities, they could hold that identity can be indeterminate (since they could allow that there might be cases where it is neither true nor false that an original Ego persisted). In order to hold that identity cannot be indeterminate, the Cartesian must therefore reject ontic vagueness of Egos.

What Parfit fails to recognise here is that someone might think that identity is always determinate whilst denying that we are separately existing entities *precisely because* they think that identity does not consist in anything else. The primitivist position we are presently considering rejects indeterminacy. Because of that, our primitivist believes that there is a sharp cutoff on the spectrum. But being a primitivist, they hold that facts about persistence do not consist in any other facts. So they deny that the difference between identity and non-identity is grounded by facts about Cartesian Egos. And so it is *because of this* that they don't need to appeal to the existence of Cartesian Egos to explain the difference between identity and non-identity. Parfit's objection to the view that there is a sharp cutoff in the absence of a Cartesian Ego is, in essence, that it can't explain the difference between identity and non-identity. But his objection clearly assumes that the difference is explainable – which is just to beg the question against the primitivist. Parfit therefore fails to show that the primitivist who rejects (SP2) is in an indefensible position.

There is another consideration that might appear to count against the primitivist who rejects (SP2): the problem of proportionality. Intuitively, it seems that something of great significance (a person's survival) cannot turn on something of great trivialness (the difference of a subatomic particle or suchlike):

> We are inclined to believe that there is *always* a difference between some future person's being me, and his being someone else. And we are inclined to believe that this is a *deep* difference. But between neighbouring cases in this Spectrum the differences are trivial. It is therefore hard to believe that, in one of these cases, the resulting person would quite straightforwardly be me, and that, in the next case, he would quite straightforwardly be someone else. (Parfit 1987: 239)

It seems that the primitivist who believes in a sharp cutoff on the spectrum is committed to the view that a trivial empirical difference will, as Parfit (1987: 239) says, 'make all the difference' between identity and non-identity. Since the difference between identity and non-identity is a matter of great significance, the primitivist must be mistaken.

This isn't quite right, though. We need to get clear about what the primitivist believes. Strictly speaking, the primitivist does not believe that a trivial empirical difference will 'make all the difference' between identity and non-identity. The primitivist doesn't think that anything *makes* it the case that identity is preserved or lost. The facts of persistence and cessation, when they obtain, are brute. And so if we are to say that the primitivist holds that identity 'turns on' a trivial empirical difference, we must understand 'turns on'

not as ‘consists in’ but as ‘coincides with’. And when this is made explicit, the objection loses its force. Parfit’s claim is that a trivial difference cannot *constitute* a significant difference. We can see that, with regards to personal identity, the primitivist would in fact agree with Parfit here – they think that no difference *whatsoever* can constitute the significant difference between identity and non-identity. And so they are immune from Parfit’s objection.

Perhaps a stronger, positive proportionality claim might be suggested. Not only must it *not* be the case that a significant difference be constituted by a trivial difference; it *must* be the case that whatever constitutes a significant difference is itself of proportional significance. This principle is also ineffective against the primitivist, however. For the primitivist, the difference between identity and non-identity consists in nothing other than itself. In other words, they think that the difference between persistence and cessation consists *only in itself*. So on their view the significant difference *does* consist in a significant difference – itself! If the principle is to be effective against the primitivist, it must be amended as: a significant difference must be constituted by something *other than itself* that is of proportional significance. But this is, again, to beg the question against the primitivist. The problem of proportionality is no problem for the primitivist. The primitivist is free to reject (SP2). The Shoemaker-Parfit argument is ineffective against the primitivist. So far, so good for the Further Fact View.

1.4.2 Criterialist arguments against the Further Fact View

If we are to find arguments against the primitivist’s Further Fact View, we would do better to look elsewhere. ‘Primitivism’ is not a term one finds being used in the literature on personal identity. As we saw, a primitivist fails to say anything informative whatsoever about what identity consists in, and so they reject R1 – and, likewise, C1. Since Criterialists are committed to C1 and C2, this means that the primitivist rejects Criterialism. If we are looking for arguments against the Further Fact View, it might be a good idea to take a look at the arguments advanced by Criterialists against Anti-Criterialism.

The standard objections against the Anti-Criterialist are epistemological ones, focusing on their denial of necessary and sufficient conditions for persistence. One unsuccessful line of objection focuses on the lack of *necessary* conditions. If there are no necessary conditions for persistence, the thought goes, there is nothing stopping personal identity

from swapping between entities that are discontinuous in every empirical (observable and experienceable) regard. The swap might be between distinct human beings, or even between a human being and, say, a rock. The fact that the entities are discontinuous in every empirical regard is irrelevant; no relation anyone could possibly observe or experience could rule the identity-swap possibility out. Sydney Shoemaker must have something like this in mind when he objects that the Anti-Criterialist 'seems committed to holding that a series of property instances whose members are causally unrelated could nevertheless be the career of a persisting object' (2012: 126). But the Anti-Criterialist is by no means committed to such a view. As Merricks points out (1998), Anti-Criterialism is perfectly consistent with the view that there are necessary conditions for persistence. So in response to Shoemaker, an Anti-Criterialist could just allow that causal continuity is necessary for persistence, but not sufficient for it. And by doing so, their view would guarantee that identity only ever holds between object-stages that are causally continuous – thus overcoming the sort of concern voiced by Shoemaker.

Another line of objection focuses on the Anti-Criterialist's lack of *sufficient* conditions (Duncan 2014; Shoemaker 2012; Olson 2012; Zimmerman 1998). Granted, in positing necessary conditions the Anti-Criterialist is able to guarantee that two persons *cannot* be identical unless they are empirically related in certain ways. But because none of those empirical relations are sufficient for personal identity, the Anti-Criterialist cannot ever guarantee that those two persons *are* in fact identical. Even if physical continuity is necessary for personal persistence, we could have diachronically physically connected and continuous persons who were nevertheless numerically distinct. Indeed, because physical continuity (or whatever empirical relation we are supposing to be a necessary condition for persistence) would not *guarantee* persistence, the Anti-Criterialist has to allow for the possibility that persons *never* persist. Nothing we could ever observe could count as evidence against that possibility.

1.4.3 Refining the primitivist's position

Can we now conclude that Anti-Criterialism, and therefore the primitivist's Further Fact View, is done for? Well, not quite. It depends on the relation between C1 and C2. Recall that the Criterialist holds that C1 (identity consists in other facts) and C2 (there are necessary and sufficient conditions for identity) are true. Taking that as our definition, an Anti-Criterialist will then be someone who thinks that *at least* one of those claims is false. The preceding line of argument counts against someone who thinks that C2 is false. But

this will only count against Anti-Criterialism *per se* if rejecting C1 commits one to rejecting C2. If it is possible to reject C1 but *affirm* C2, the Anti-Criterialist will be able to avoid the objections above.

What is the relation between C1 and C2 then? There is, I think, something intuitive about the idea that C1 and C2 are biconditionally related. The intuition might go something like this. If on the one hand you hold that the relation of identity does *not* consist in any other relation, then surely the most you can say is that *identity* is necessary and sufficient for persistence. You won't be able to supply any non-trivial necessary and sufficient conditions – in which case not-C1 entails not-C2; and so C2 entails C1. This appears to be the sort of reasoning Sydney Shoemaker is employing when he says:

> The diachronic unity relation cannot be simple and unanalysable if its obtaining requires spatio-temporal continuity or connectedness. (Shoemaker 2012: 129)

As I understand it, to say the diachronic unity relation is simple and unanalysable is to say that it does not consist in anything other than itself. If that's right, then Shoemaker is saying that someone who thinks that spatio-temporal continuity is necessary for persistence must also think that persistence consists in something other than itself. So by Shoemaker's lights, if there are necessary conditions for persistence, then persistence will consist in something other than itself. That is, Shoemaker thinks that C2 entails C1.

If on the other hand you hold that the fact of a person's persistence consists in other facts, then surely you will also hold that the obtaining of those other facts will be necessary and sufficient for that person's persistence. For example, if you think that one's persistence consists in the existence of a series of persons with whom one is physically continuous and/or connected, it might seem to follow that you will hold physical continuity to be necessary and sufficient for persistence. If that's the case, C1 entails C2. Intuitively, then, there is a case for supposing that C1 and C2 are biconditionally related. I take it that this intuition, or something like it, is demonstrated by the tendency to lump C1 and C2 together; and that it is further demonstrated by the Criterialists' apparent assumption that refuting Anti-Criterialism can be achieved simply by showing why C2 is indispensable.

This intuition should not be trusted, however. For starters, as Olson (2012) warns, '[p]erhaps our identity over time could consist in something that is not expressible in terms of necessary and sufficient conditions' (2012: 49). The suggestion here is that C1

could be true and yet C2 be false, casting doubt on the conditional C1→C2. I do not want to spend any time considering how this might work; the relation we are really interested in is not the biconditional itself; what we particularly want to know is whether the conditional C2→C1 is true. If this conditional is false, the Anti-Criterialist will be able to reject Criterialism by accepting C2 but rejecting C1; and, in accepting C2, will be immune to the Criterialist's epistemological objections considered above.

In saying that persistence does not consist in anything else, one is saying that, whenever it is a fact that a person persists, this fact is not made true by any other facts. That the person persists is a 'brute' or 'primitive' fact. Now, there does not appear to be any *contradiction* in the thought that, even though this fact is brute, it nevertheless always happens to be accompanied by certain other facts, such as the holding of particular empirical relations. This would amount to the view that personal persistence is brute but that it has necessary conditions attached. And further, there does not seem to be any *contradiction* in the thought that, in addition, there are no instances when these other facts obtain but the brute fact of persistence doesn't. This would mean that these other facts are sufficient for persistence. Thus, someone who believed all these things would be a primitivist who believes there to be necessary and sufficient conditions for persistence. There is logical space for someone to reject C1 but accept C2.

Indeed, the reason that Criterialists need to insist on C1 in addition to C2 is that philosophers are not simply interested in what evidence we use to *judge* that personal identity obtains; they want to know what *makes it the case* that personal identity obtains (Olson 2012). If we knew what the necessary and sufficient conditions for personal identity were, we would be in a position to make identity judgements. If we knew that, say, physical continuity was necessary and sufficient for persistence, we would know that whenever physical continuity doesn't obtain neither does identity; and whenever physical continuity does obtain, so identity does too. So we would be able to say *that* the person we are observing now is the same person we were observing some time ago. But we wouldn't merely in virtue of that be able to say *why* the person we are now observing is the same person we were observing before. This is because the biconditional contained in C2 is symmetrical: if physical continuity (for example) is necessary and sufficient for personal identity, then personal identity is also necessary and sufficient for physical continuity. Hence, C2 could be true and it nevertheless be the case that personal identity does *not* consist in physical continuity; it might instead be the case that physical

continuity consists in personal identity. Or perhaps personal identity and physical continuity aren't related by a grounding relation at all. Hence, C2 does not entail C1. All C2 tells us is that there is a *correlation*. We need C1 to impose an explanational asymmetry on the correlation; we need C1 in order to say what *grounds* persistence.

Since C2 does not entail C1, the epistemological objections we discussed earlier that were levelled against Anti-Criterialism are not effective against Anti-Criterialism *per se*, but rather against a specific variety of Anti-Criterialism – namely, that which rejects C2. Admittedly, this Anti-Criterialist view is the one we typically find populating the literature. But if we are to follow the definition of Anti-Criterialism stipulated by the authors at the beginning of this chapter, we must recognise there is space for an Anti-Criterialist to accept C2 whilst rejecting C1; an Anti-Criterialist can be a primitivist who holds that brute facts can have necessary and sufficient conditions. This Anti-Criterialist is immune to the criticisms examined so far. That is not to say the view does not face other difficulties – we will get to those in the following sections. But it just means that the Criterialist's objections considered so far are only half the story.

From here on in, I will be assuming that the primitivist accepts the existence of necessary and sufficient conditions for persistence. For ease of exposition, then, I will use the term 'primitivist' to refer to the Anti-Criterialist who denies C1 but accepts C2. There are two arguments I want to make against such a view. Firstly is a general concern about theoretical simplicity. Secondly is the question of why we should believe that identity has necessary and sufficient conditions if it consists in no other facts.

1.4.4 Arguments against the refined primitivist position

The general concern we should have about primitivism is that, all other things being equal, primitivism posits more unexplainable facts. This is a point Lewis and Sider make against modal primitivism: 'primitive modality is bad news, and more kinds are worse than fewer' (Lewis 1986: 242; see also Sider 2011: 267). The primitivist about persistence claims that the fact of a person's persistence does not consist in facts about person-stages and their relations; the existence of a persisting person is a *further* fact. For the informativist on the other hand, the existence of a persisting person is *not* a further fact and as such theirs is the simpler theory. They are able to explain more, and with fewer brute facts. All other things being equal, then, we should prefer informativism as the simpler, more epistemically responsible theory.

It might be contended by the primitivist that *theirs* is in fact the simpler theory. After all, they say that persistence does not consist in *anything but itself*, whereas the informativist says that persistence consists in all these *other* facts. Indeed, primitivism about personal identity is sometimes referred to as 'the simple view'.

This response from the primitivist misses the point of the objection. The primitivist can accept that the persisting person is intra-related in certain ways. They do not deny that these facts are *involved* in the persistence of a person. And if they think there are necessary and sufficient conditions for persistence (as they will on our current definition of the term 'primitivist'), then these facts are *necessarily* involved. Thus, they will agree with the informativist that these finer-grained facts obtain. The difference is that the informativist thinks that *there is nothing more to a person's persistence than these finer-grained facts*. And so, if the informativist were to write an inventory of all the facts involved in a person's persistence, they would not need to include the fact of persistence itself. That fact would have already been included in the list of finer-grained facts. This is where the primitivist disagrees. The primitivist's list *would* include the fact of persistence. So the primitivist is committed to all the facts the informativist is, but with the *additional* further fact of a person's identity.

That's not all, though. Our primitivist thinks there are necessary and sufficient conditions for persistence. They think that persistence always coincides with some particular empirical relation. This, I suggest, should strike us as peculiar. There are, I think, two questions that we should ask, one metaphysical and one epistemological. Firstly, if persistence doesn't consist in that relation (or relations), what could possibly *make it the case* that persistence always coincides with a particular empirical relation (or relations)? And secondly, why should we *believe* that persistence always coincides with that relation despite not consisting in it?

With regards to the first question – *why* there is such a coincidence – the primitivist has no explanation as far as I can tell. *There just is.* Now, this in itself might be enough to convince some that primitivism is a hopeless case. But the primitivist can just say that this is another brute fact. They've already bitten the bullet with the brute fact of persistence; entertaining another brute fact might not seem like much to worry about. And we can't object to primitivism simply for positing brute facts about personal persistence. What we *can* object to, however, is the relative complexity of their view. Although it might not seem like much, positing another brute fact adds to this complexity, making the

case against primitivism stronger. We can object to the epistemic irresponsibility of a theory that multiplies the number of brute facts about the world just to keep itself afloat.

Now for the second question: why should we *believe* that identity has necessary and sufficient conditions, even though it does not consist in anything other than itself? An answer to this question has been suggested by Simon Langford (2017). His move is, in short, to say that primitivists should accept the existence of necessary and sufficient conditions *in the actual world*, even though they should deny the existence of necessary and sufficient conditions in *every possible world*. Whilst they should allow that it is *possible* for identity to float free from, say, physical continuity, they should accept that it always coincides with the relation whenever *we* observe it:

> [T]here are possible worlds where people bearing all the qualitative relations that ordinarily accompany our persistence are distinct. Nonetheless, it's also possible on this view that at the actual world a contingent law holds which ensures people bearing those relations are in fact identical. (Langford 2017: 62)

What the primitivist needs to posit, then, is a contingent 'gluing law' that just so happens to glue identity to (e.g.) physical continuity in the actual world. With such a law in place, it would be guaranteed that whenever we observe physical continuity (between persons) we have an instance of personal persistence. Langford's contention is that the hypothesis – that the gluing law holds in the actual world – should be believed for non-empirical reasons. His thought is that once we weigh up all the non-empirical virtues of the hypothesis we will find it more compelling than not. I do not wish to go through the list of virtues one by one and assess whether they are indeed compelling. It is the very suggestion that we should find the hypothesis compelling for non-empirical reasons that should, I think, be setting off alarm bells.

The position being assessed here allows that there are two possible worlds exactly similar in every respect, except that in one world the gluing law holds and in the other it doesn't. But if it is reasonable to believe that the gluing law holds for non-empirical reasons, the inhabitants of one world will have just as much reason to believe that the law holds in their own world as the inhabitants of the other world will have to believe it holds in theirs. So whatever the non-empirical virtues of the hypothesis, we have to acknowledge that there would be no way of knowing whether the gluing law obtains. We have to acknowledge that we could well be inhabiting the world in which it doesn't obtain. But

now we seem to have gone full circle: it's not clear how this position can avoid the epistemological problems it set out to avoid. Recall the criticism we encountered earlier against Anti-Criterialists who deny there are sufficient conditions for persistence. The problem was that no observation could guarantee that the people we intuitively identify were in fact the same person. We would have to allow for the possibility that persons *never* persist. The same can be said of Langford's proposal. We must, on Langford's proposal, recognise that whatever the non-empirical virtues of believing that the gluing law holds, we might inhabit the world in which it doesn't. We would have no way of knowing whether our observations of any particular empirical relation coincided with personal identity. And because of this we would have no way of knowing whether persons ever persist; I might not be the same person from moment to moment. It does not appear that Langford's possible world approach is going to be particularly helpful getting the Anti-Criterialist over the epistemological hurdles they face. I conclude that primitivism about personal identity should be disregarded as epistemically bankrupt. The view that personal identity is a 'further fact' is indefensible.

1.5 Primitivism and brute persistence views

We began by considering Parfit's argument against the existence of Cartesian Egos. That argument knocked out the NE-informativist$^{\mathbf{E}}$ positions, which was the middle row in Figure 3. I have just argued that primitivism must also be rejected, which knocks out the bottom-right box. My claim is that the considerations that count against primitivism will also count against the E-informative$^{\mathbf{E}}$/R-uninformative views (top right box). The views that fall into this camp are those claiming that personal identity consists in the brute persistence of empirical entities, such as brains, or bodies, or streams of consciousness.

The reason for this classification is as follows. A primitivist thinks that the fact of a person's persistence does not consist in any other facts. The E-informativist$^{\mathbf{E}}$/R-uninformativist thinks that the fact of a person's persistence does consist in another fact, and this fact is the fact of, say, a brain's persistence. But they cannot think there is anything informative to be said about the relation that grounds brain persistence. If there was something informative to be said about the relation that grounds brain persistence, then, given that they think that brain persistence grounds personal persistence, they would have something informative to say about the relation that grounds personal persistence. But then they wouldn't be an E-informativist$^{\mathbf{E}}$/R-uninformativist. So for the E-

informativist^{E}/R-uninformativist, brains just persist. There is no other relation that grounds their persistence.

That is not to say they can't say something informative about the entity that grounds brain persistence, though. They could say that a brain is a collection of particles arranged in a certain way, and it is the persistence of this arrangement of particles that grounds brain persistence. But, for the reason given above, they cannot say anything informative about the grounding relation.

So then, our E-informativist^{E}/R-uninformativist thinks that a person's persistence is grounded by the persistence of a brain, and that the brain's persistence is grounded by the persistence of a certain collection of particles. We may now ask: what does the persistence of that collection of particles consist in? Perhaps they will have an answer even for that question. But sooner or later the E-informativist^{E}/R-uninformativist will run out of answers. And when they do, they will have nothing informative to say about either the entity or the relation that, ultimately, grounds personal persistence. That is, they will be primitivists about the persistence of entities of a certain sort. And, as such, they will be prone to the same criticisms as the primitivist about personal persistence.

The same considerations that count against the primitivist also therefore count against the E-informativist^{E}/uninformativist^{R}. If the primitivist's Further Fact View is indefensible, then so is E-informativism^{E}/R-uninformativism. I therefore propose that E-informativism^{E}/R-uninformativism should be regarded as a version of Non-Reductionism even though Parfit fails to identify it as such. The only remaining position untouched by the arguments against Non-Reductionism is E-informativism^{E}/R-informativism^{E}. This, I submit, is what Reductionists hold. Thus, I propose we define Reductionism as the view that:

> R3: Facts about a person's identity just consist in facts about empirical relations holding between empirical entities.

A view is Non-Reductionist if it rejects R3. That is, Reductionist accounts are E-informative^{E}/R-informative^{E}, whereas Non-Reductionist accounts are either E-informativist^{NE} (Cartesian Views) or R-uninformativist (which include the Further Fact View and the other brute persistence views).

1.6 Summary

The difference between Reductionism and Non-Reductionism does not turn merely on whether or not identity is taken to consist in other facts, as Parfit sometimes claims. It matters what these other facts are. Criterialists and Reductionists agree on C1; they both agree that personal identity consists in other facts. But whereas the Reductionist stipulates what *sort* of facts personal identity consists in (facts about empirical relations holding between empirical entities), the Criterialist makes no such stipulation. Since all Non-Reductionists except for primitivists hold that identity consists in other facts, all Non-Reductionists except for primitivists can be Anti-Criterialists (provided they think that the facts in which identity consists are necessary and sufficient for identity – satisfying C2).

The arguments considered have shown that any view that is not E-informativist$^{\mathbf{E}}$/R-informativist$^{\mathbf{E}}$ should be rejected. We first considered Parfit's arguments against Cartesian Egos, which were effective against the E-informativist$^{\mathbf{NE}}$ positions. The prospects for the primitivist's Further Fact View were then assessed. Shoemaker's argument against the Further Fact View was found to be ineffective, even when bolstered by considerations against primitivism found in Parfit's work. The efficacy of arguments made by Criterialists were then considered. To avoid the Criterialist's standard criticisms against Anti-Criterialism – that we could never know if identity held between persons – the primitivist must hold that there are necessary and sufficient conditions for persistence. But doing so lands them in more hot water. Firstly, primitivism loses out to informativism in terms of simplicity and explanatory potency – a matter amplified by the primitivist's reliance on necessary and sufficient conditions. Secondly, we would need to posit a 'gluing law' as suggested by Langford that 'glues' identity to a chosen empirical relation (such as physical C/C). But we could never know whether such a law actually obtains. And so, again, we could never know if identity held. It was concluded that the primitivist's Further Fact View is untenable. Finally, I argued that the considerations against primitivism also count against the brute persistence views: E-informativism$^{\mathbf{E}}$/R-uninformativism. Throughout all this, one group of views is left untouched: those that say something empirically informative about both the entity and the relation. This, I propose, is what we should take as our definition of Reductionism. All other views should be rejected.

Chapter 2: Reductionism and Empty Questions

The aim of this chapter is to establish that, given the truth of Reductionism, it is an empty question whether persons, and their experiences, persist through time. Parfit uses his Combined Spectrum thought experiment to argue that in some cases, questions of personal identity could be empty. I contend that this conclusion is not radical enough. After examining the Combined Spectrum, I argue that a stronger conclusion is warranted: questions of personal identity are *always* empty. I then go on to argue that because of this, questions of experiential identity are empty too. I begin by explaining what it is for a question to be 'empty'.

2.1 Empty questions

A question is empty, according to Parfit's terminology, when different answers to that question are merely different ways of describing the situation in question:

> Suppose a certain club exists for several years, holding regular meetings. The meetings then cease. Some years later, some of the members of this club form a club with the same name, and the same rules. We ask: 'Have these people reconvened the *very same* club? Or have they merely started up *another* club, which is exactly similar? …. We would not be puzzled when we cannot answer the question, 'Is this the very same club?' We would not be puzzled because, even without answering this question, we can know everything about what happened. If this is true of some question, I call this question *empty*. When we ask an empty question, there is only one fact or outcome that we are considering. Different answers to our question are merely different descriptions of this fact or outcome. This is why, without answering this empty question, we can know everything that there is to know. (Parfit 1987: 213-4)

When people disagree over how to answer an empty question, their disagreement can be regarded as merely verbal. They are not disagreeing about the state of the world, because they already 'know everything there is to know'. Rather, the difference between their answers simply reflects the difference in the language they use to *describe* the world.

Parfit also claims that answers to empty questions have indeterminate truth values. He thinks that if we answer the above question by saying that the two clubs are indeed identical, this answer will be neither true nor false (1987: 213). It is not immediately clear why this should be the case. To illuminate matters, it will help to take a look at Ted Sider's (2011) discussion of 'merely verbal' disputes. Merely verbal disputes arise, says Sider, when we try to answer what he calls 'nonsubstantive' questions – for example, the question of whether Robinson Crusoe is a bachelor:

> [F]or one or more expressions E (e.g. 'bachelor') in a nonsubstantive question, the semantic candidates for E (unmarried-adult-male, unmarried-adult-male-eligible-for-marriage, etc.) are such that i) each opposing view about the question comes out true on some candidate; and ii) no candidate carves at the joints [of nature] better than the others. If E uniquely means one of these candidates, this is not because that candidate is intrinsically privileged. It is only because our linguistic community happened to select that candidate rather than one of its mates as the meaning of E. (Sider 2011: 46)

Applying this line of thought to Parfit's club example, it is clear that the expression 'same club' has multiple semantic candidates.[1] For one candidate, 'same club' might be satisfied so long as the same name is used and at least some original members remain. For another candidate, it might be necessary for *all* members to remain, such that a club cannot retain its identity if there is a change in any of its members. Neither of these candidates, it seems plausible to suppose, 'carves nature at its joints' better than the other. Whether or not linguistic convention favours one over the other, neither seems to be more metaphysically accurate. So even if we disagreed with someone over whether the club retained its identity, it would be a mistake to think that they were making a mistake about the way the world is. They would merely be employing a different semantic candidate in

[1] There is the question of what qualifies as a semantic candidate. Sider allows that there will be a degree of 'mismatch with usage' (2011: 50), in that an alternative candidate might not match our usage of a term. But at the same time, the mismatch can't be too jarring: 'If a linguistic community, roughly in our circumstances, *could* have used E to mean m without seeming "semantically alien"—could have used E to reach "the same semantic goal" as we use E to reach, albeit perhaps by a different route—then m is a candidate for E.' (*ibid.*). The conclusion I am ultimately arguing for is revisionist: we *could* describe persons as non-persisting entities. My contention is that many of our 'semantic goals' could be still be achieved if we were to describe the world in this way. But, as a B-theorist and a Reductionist, I hold that some of our semantic goals – those concerning the expression of A-theoretic ideas and certain self-directed attitudes – should be abandoned. I discuss this more in chapter 5, when I compare our conceptual scheme with a possible alternative based on this different way of speaking.

answering the question. The question is empty; the disagreement would not be substantive.

Although Sider's characterisation of a nonsubstantive question echoes that of Parfit's empty question, there is a *prima facie* tension between the two: whereas Sider says that *both* answers will be true, Parfit says that *neither* answer can be true (or false). We needn't understand these two views as contradictory, though. Sider says that both answers are true, but only *relative to their respective semantic candidates*. When an answer is true in this way, we can say that it is *nonsubstantively* true.

Nonsubstantively true: true relative to the semantic candidate employed, where there is no privileged semantic candidate.

In the same vein, we can see that there is a sense in which both answers will be nonsubstantively *false*: although both answers come out true relative to their *respective* semantic candidates, they will come out false relative to the *opposing* candidate.

Nonsubstantively false: false relative to an opposing semantic candidate, where there is no privileged semantic candidate.

We can therefore say that in the club example both answers will be nonsubstantively true *and* nonsubstantively false. Hence, neither answer will be *substantively* true or false. This, I take it, is why Parfit claims that answers to empty questions have indeterminate truth values. Neither answer can tell us anything about the world; they can only tell us about how the utterer *describes* the world. So if we want to know about the way the world is, and we want to assign truth values to claims on the basis of what they tell us about the world, we will have to remain silent. Neither answer is true or false in the substantive sense.

We have all overheard – and no doubt regrettably participated in – the sort of tiresome verbal dispute that ensues when an empty question isn't recognised for what it is. And it is easy to see that any disagreement over the question in Parfit's club example would be one such instance. But it is not at all obvious that we should say the same about questions of personal identity.

Intuitively, my continued existence is very different from the continued existence of a club. I can remember waking up this morning, sitting at my desk, writing the start of this sentence, and so on. If I asked myself: 'am I the same person who began writing this sentence?' I would instinctively answer 'yes'. And in doing so I would be expressing what seems to be a 'deep' truth – a truth about the way the world is. If someone were to disagree with me, I would think they were confused. If they were to claim in all seriousness that the person who started writing this sentence was a *different* person – not just qualitatively but *numerically* different – and that I – me, now – had just suddenly popped into existence, inhabiting what used to be their body and remembering their experiences – I would be worried for their mental health.[2] The disagreement between us would seem to be of a very different sort from the disagreement in the club example; our disagreement seems substantive. And it is hard to imagine that this sort of disagreement could *ever* be nonsubstantive. According to common sense,

> We are inclined to believe that there is *always* a difference between some future person's being me, and his being someone else. And we are inclined to believe that this is a *deep* difference. (Parfit 1987: 239)

For any person at any time, the intuition goes, there will be a yes or no answer to the question of whether that person is me. Any disagreement over whether or not identity holds will be substantive. Unlike the identity of clubs, there *is* a best way to carve nature at its joints – and given this superior carving *all* answers to questions of personal identity will be either true of false. Or so it seems.

Parfit argues that this is wrong: there are some conceivable cases in which there is no 'deep' difference between my being identical to some other person or not, and that disagreements over this matter will be merely verbal. Questions of personal identity can be empty. We can now turn our attention to the Combined Spectrum argument. I will outline Parfit's argument first, before identifying gaps in need of shoring up.

[2] A matter of terminology: one might object that it would be impossible to *remember* someone else's experiences, since strictly speaking one can only remember *one's own* past experiences. This can be avoided easily by introducing the notion of a 'quasi-memory' or q-memory, which does not require personal identity.

2.2 Parfit's Combined Spectrum Argument

To recall, the Combined Spectrum is an imagined range of operations in which the empirical properties of person p1 (Derek Parfit) are replaced with those of person p2 (Greta Garbo). At the near end of the spectrum, some minutiae of p1's empirical properties are replaced with those of p2. In the next case along, a trivially larger amount of p1's empirical properties are replaced with those of p2. This continues until we reach the far end of the spectrum, where we have an operation in which *all* of p1's empirical properties are replaced with those of p2.[3]

It is common sense that persons survive minor physical and psychological changes. If, while I am clipping my fingernails I forget where I put my keys, I do not cease to exist, replaced by an almost-exact replica. So common sense dictates that the first few operations on the spectrum should be regarded as ones in which p1 retains their identity. Thinking now about the far end of the spectrum, it is equally common sense that a person will die if their brain and body are destroyed, and their psychological traits eradicated. So at the far end of the spectrum, where we are to suppose that all of Derek Parfit's distinctive physical and psychological features are replaced by those of Greta Garbo, we have a clear case of non-identity.

If there is always a 'deep' difference between identity and non-identity, then, for any operation on the spectrum, there will be a yes or no answer to the question of whether p1 retains their identity. Towards the near end of the spectrum the answer will be 'yes', and towards the far end the answer will be 'no'; and somewhere on the spectrum there will be a sharp borderline separating the two regions. Parfit argues that such a view is indefensible (1987: 239).[4]

Firstly, there can be no evidence of such a borderline. If there was a borderline it would lie between a specific pair of adjacent cases. But the difference between any pair of

[3] Parfit's Combined Spectrum thought experiment is an analogue of Chisholm's (1967). Chisholm is interested in trans-world identity – that is, what makes person p1 in possible world W1 the same person as p2 in W2. The modified Parfittian argument I develop in this chapter might, with some alterations, be applied to the possible worlds question.

[4] The Combined Spectrum can be regarded as laying the groundwork for two related arguments. One is an argument purporting to show that we should reject Non-Reductionism, and the other is an argument that, assuming the truth of Reductionism, shows that identity can be determinate (and that there is not always a 'deep difference' between identity and non-identity). Having argued for Reductionism in the previous chapter, I focus in this chapter on the second of these two arguments.

adjacent cases on the spectrum is just as trivial as any other pair. As such, there would be no empirical reason to favour one pair of cases over another as the borderline's location. No observation could support the claim that in one particular operation p1 survives, whilst in the next they don't. Call this the *evidence problem.*

Secondly, we intuitively believe that there is always a deep difference between identity and non-identity. But positing a sharp borderline entails that this 'deep' difference will turn on some trivial physical and/or psychological change. If you believed there was a sharp borderline, you would have to think that a different cell (or molecule, or subatomic particle) and different vague memory (or a slightly stronger desire, or a slightly weaker conviction) would make the all the difference between life and death. But, Parfit asks,

> What could make it true that, in one case, the resulting person would be me, and in the next he would not be me? What would the difference consist in? (Parfit 1987: 239)

Call this the *truthmaker problem*. Parfit claims that only the Non-Reductionist is able to answer these concerns, and so we can only accept the existence of the sharp borderline if we also accept Non-Reductionism.[5]

Either the Non-Reductionist will be a Cartesian, or they will be a primitivist (including the brute persistence views). If the former, they can hold that in some of the cases p1's Ego remains throughout the operation, whereas in others it doesn't; and that between these two cases is where the borderline lies.[6] Given that Egos are unobservable, the Cartesian needn't be worried by the lack of evidence. So the Cartesian Non-Reductionist is untroubled by the evidence objection. And although they couldn't say *where* on the spectrum the borderline lies, they are nevertheless able to say *why* it is true that it lies between whichever cases it does – *viz.* because the original Ego remains in one case, and doesn't in the next. This, for them, is what the difference between identity and non-

[5] Parfit actually only mentions that the *Cartesian* Non-Reductionist is able to claim there is a sharp borderline (1987: 293), but, as shown, the primitivist can also make this claim.

[6] That is not to say a Cartesian *must* hold that there is a sharp borderline. They will if they think that the existence of an ego is always all-or-nothing. But they might instead think that egos are ontically vague entities (more on that later). The point is that they at least have the *option* to say that the existence of egos is always determinate; and if all-or-nothing egos did exist, the sharp borderline would sit between two trivially different adjacent cases.

identity consists in. Thus, the Cartesian Non-Reductionist also has an answer for the truthmaker objection.

The primitivist Non-Reductionist is also unruffled by the evidence and truthmaker problems. There *is* a fact of the matter about which operations p1 survives, and this fact does not consist in other facts.[7] So given that the fact of a person's identity is brute, a lack of evidence is precisely what we should expect. And so it would be missing the point to ask *what could make it true* that in one case p1 survives and in the next they don't. For the primitivist *nothing* makes it true.

The Reductionist, on the other hand, believes that the fact of a person's identity consists in empirical facts. Given that these facts are empirical, they cannot, as the Non-Reductionist does, accept the lack of evidence for p1's identity and non-identity. Nor can they avoid the truthmaker problem. Unlike the primitivist, the Reductionist must hold that the difference between identity and non-identity consists in other facts; but unlike the Cartesian, they cannot hide behind any posited unobservable facts to explain the difference. The Reductionist has nothing they can point to that would make it true that the borderline lies between one pair of cases rather than another. And so, given that Reductionism is true, we must reject the idea of a sharp borderline. Our intuitions in that regard would only be justified if Non-Reductionism were true. And so, Parfit concludes, as Reductionists we must accept that there need not always be a 'deep' difference between identity and non-identity:

> The resulting person would be me in the first few cases. In the last case he would not be me. In many of the intervening cases, neither answer would be true. I can always ask, 'Am I about to die? Will there be some person living who will be me?' But, in the cases in the middle of this Spectrum, there is no answer to this question … This question is, here, *empty*. (Parfit 1987: 232-233)

2.3 Objections and clarifications

I think Parfit's conclusion is going in the right direction. But it isn't radical enough. Parfit's halfway position is, I shall argue, unjustified. What his argument shows – once

[7] The primitivist is not *committed* to the existence of a sharp borderline. As we saw in Chapter 1, they might hold that in some cases it is neither true nor false that p1 survives, and that this fact of indeterminacy is brute. Again, the point is that a primitivist *could* be justified in believing in a sharp borderline.

we have ironed out a few crinkles – is that *all* questions of personal identity along the spectrum are empty. Before I get on to explaining why, it will be helpful to address some potential objections and points that are in need of clarification. Doing so will clear the ground in preparation for my stronger claim.

2.3.1 Proportionality

As discussed in Chapter 1, Parfit thinks that a trivial difference in empirical facts cannot make a 'deep' difference between identity and non-identity. But this, Sorensen (1988) argues, is to assume a false principle of proportionality:

Proportionality Principle: 'the magnitude of a modification must be proportional to its effect' (Sorensen 1988: 251).

Sorensen argues that we should reject the Proportionality Principle, and with it Parfit's conclusion from the Combined Spectrum. We should reject the Proportionality Principle because small changes *can* make all the difference. To lend plausibility to this claim, Sorensen considers a rocket launch. In order to overcome Earth's gravity, the rocket must reach a certain speed. But this speed constitutes a sharp borderline; just below the borderline, and the rocket will come crashing back down to Earth. As such, a single drop of rain slowing the rocket down by an infinitesimal amount is enough to make the difference between a successful launch and catastrophe. The proportionality principle is wrong.

In fact, though, the principle of proportionality that Parfit assumes isn't the one attacked by Sorensen. We can see this by comparing the rocket launch example with the Combined Spectrum. In the case of the rocket launch we have the facts about the launch condition, which differ trivially in the two cases considered. But the outcome of the launch is a further fact over and above the facts about the launch conditions. In the Combined Spectrum, the fact of the person's identity is not a further fact (whether that be a fact about an Ego or a primitive fact); given the truth of Reductionism, the fact of a person's identity *consists in* the trivially different facts. The question in the rocket launch example is whether a trivial difference would *cause* a significantly different state of affairs. But the question in the Combined Spectrum is whether a trivial difference would *constitute* a significantly different state of affairs (Unger 1990; Alter & Rachels 2005). Sorensen's

proportionality principle is about differences in causes and effects. This misses the point of the Combined Spectrum argument, and the role that Reductionism plays in it.

2.3.2 Animalism

Because it alters physical and psychological properties, the Combined Spectrum argument is effective for Reductionists who espouse either the physical or psychological criteria of personal identity. However, there is a criterion that doesn't fit neatly into either category: animalism. According to animalists, 'each of us is numerically identical with an animal' (Olson 2003: 318).[8] It is not out of the question that an animalist might hold a primitivist view: that a person's identity consists in the brute persistence of a human animal (i.e. where that animal's persistence consists in no other facts). They would then be a Non-Reductionist, and so would be justified if they were to maintain that the spectrum has a sharp borderline.

However, the common animalist view is an empirically informative one.[9] That is, the fact of a person's existence consists in empirical facts that pertain to the continued existence of an animal – whatever those empirical facts might be (views differ). And so most animalists will be Reductionists. However, the *persistence* conditions of animals and those of persons are often held to be different. For example, it is common to regard an early foetus as an animal but not a person. Equally, some animalists think that animals persist as corpses (e.g. Carter 1999; Mackie 1999; Feldman 2000), but it would be a much stronger and more controversial claim to say that *persons* survive death. Nevertheless, for the informativist animalist the persistence conditions of persons will be stated in terms of empirical facts (whatever those facts may be). Accordingly, in regards to the Combined Spectrum the question we are interested in is: at what point should the animalist regard the animal *qua* person to lose its identity? And to answer this question the animalist need only consider the distinctness of animals. As Parfit puts it:

> We are asking whether, if some part of an animal's body were destroyed and replaced, the resulting animal should be counted as one and the same. This is not a scientific question [but] a purely conceptual question. (Parfit 2008: 200)

[8] The doctrine is difficult to define precisely (see for example Johansson 2007). For our purposes, however, this quick gloss will do.

[9] All references to animalism shall refer to the empirically informativist view, unless otherwise stated.

The animalist can therefore think of the Combined Spectrum as a question about transplants: in the first cases, it is intuitive to suppose that the original animal is receiving a transplant. In the last cases it seems intuitive to say that what is left of the original animal is being transplanted into a new, distinct animal. But at what point does the recipient organism become the donor part?

Olson's view is that only the brainstem is of vital importance, his reason being that it controls much of the functioning of the organism (1997; 2003). Any other organ, such as the liver, could be removed and replaced without changing the identity of the organism. The resulting animal would be the same animal, just with a transplanted liver. Even the cerebrum is expendable; all but the brainstem could be replaced, and we would still have the original (albeit drastically altered) organism. But as soon as a different brainstem is introduced, we are looking at a new animal.

As Belshaw (2011) notes, when we look closely at Olson's reasoning, he assumes that the replacement of a brainstem would result in a period of 'metabolic anarchy' (Olson 2007: 140). And it is because of this gap in regulated functioning that he believes the original organism would cease to exist. But there is no need to posit such a gap. This is science fiction, after all. Belshaw imagines the body being hooked up to a 'brainstem machine' that regulates metabolic functioning throughout the operation. Equally, we might imagine the transplant happens one nerve at a time, or that it happens via instantaneous teleportation. The point is that we don't need to suppose brainstem replacement is necessarily accompanied by a period of 'metabolic anarchy'.

The question about animal identity is not therefore about metabolic functioning; it is, in Olson's case at least, a question of brainstem identity. And so Olson's criterion of animal identity is just as susceptible to the Combined Spectrum as any physicalist criterion of personal identity: how much of the brainstem can be replaced before it becomes a different brainstem? More generally, whatever bodily parts the animalist thinks are of pivotal importance in the identity of animals, the lessons of the Combined Spectrum will apply. In holding an empirically informativist view they are prone to the same evidence and truthmaker objections that the Reductionist ran into above.

2.3.3 Vagueness

The Combined Spectrum is a version of the Sorites puzzle. In the classic Sorites, the puzzle concerns heaps. We start off with a heap of sand, and in a series of trivial changes

– the removal of a single grain – we get to the absurd conclusion that a pile of no grains is a heap. We get to this conclusion because at no point along the series is there a pair of adjacent cases where on one side we clearly have a heap, and on the other side we clearly don't have a heap. At no point does it seem reasonable to posit a sharp borderline.

Parfit argues that we should conclude from the Combined Spectrum that identity can be indeterminate. But he asserts that this indeterminacy is due to the emptiness of identity questions:

> I need not solve the Sorites Problem. It will be enough to make the following remarks. […] [W]hen the Sorites Argument is applied to heaps, we are happy to solve the problem with a *stipulation*: an arbitrary decision about how to use the word 'heap'. […] We do not believe that any collection of sand must either be, or not be, a heap. We know there are borderline cases, where there is no obvious answer to be question 'Is this still a heap?' But we do not believe that, in these cases, there must *be* an answer, which must be either Yes or No. We believe that, in these cases, this is an empty question. Even without answering the question, we know everything. (Parfit 1987: 232-3)

We should take this assertion with a pinch or two of salt. Indeterminacy is, in Sorites puzzles at least, widely associated with, or held to arise as a result of, vagueness – a topic which Parfit skirts around.[10] And it is not at all clear that vagueness in the Combined Spectrum cases should entitle us to conclude that the identity questions are empty. Without an account of vagueness, Parfit's reliance on the Combined Spectrum is susceptible to criticism (see e.g. Goodenough 1996).

To get an idea of how vagueness might fit into Parfit's argument, we can reintroduce the notions of substantive and nonsubstantive truths. To recall, Parfit's target in the Combined Spectrum is the intuition that 'there is *always* a difference between some future person's being me, and his being someone else [and] that this is a *deep* difference' (Parfit 1987: 239). Let's unpack this a second. For there to always be a difference between identity and non-identity, it must always be the case that we can answer questions of identity (does p1 retain their identity?) with a 'yes' or a 'no'; questions of identity will always have determinate answers. For there to be a 'deep' difference between identity

[10] Vagueness is also associated with *higher-order* indeterminacy, where there are borderline cases of borderline cases (see e.g. Fine 1975; Sainsbury 1991; Wright 1992; Williamson 1999a). As we will see later, this is something Parfit struggles with in his treatment of the Combined Spectrum.

and non-identity, disagreements cannot be merely verbal. There must be a correct (or at least, uniquely superior) way of carving nature at its joints which makes some answers substantively true and other answers substantively false. We can therefore regard Parfit's target intuition as having two components:

(a) All questions of personal identity are substantive.
(b) All questions of personal identity have determinate answers.

If (a) and (b) are both true, there will be a single sharp borderline on the spectrum representing the point at which it ceases to be substantively true that p1 survives and becomes substantively false that p1 survives. We can call this borderline a *substantive borderline*. Parfit's argument, if successful, would show that there cannot be a substantive borderline on the spectrum. And so it would show that (a) and (b) cannot both be true. But that leaves open the following possibilities.

(i) (a) is false and (b) is true
(ii) (a) and (b) are false
(iii) (a) is true and (b) is false

(i) is the most straightforward (but also untenable, as will become clear in discussion of (ii)). If (a) is false, questions of identity will (at least in some cases) have nonsubstantively true answers. The answers will be true relative to the semantic candidates they employ, while none of the potential candidates pick out a metaphysically privileged entity. As such, disagreements will be merely verbal. Parfit's conclusion – that questions of identity can be empty – would be secured.

If (b) is true in addition to (a) being false, identity questions will have nonsubstantively true determinate answers. That is to say, the question of whether p1 survives will have either a 'yes' or a 'no' answer, but these answers will only be nonsubstantively true. We can spell out what this would mean on the spectrum.

Suppose it was nonsubstantively true that p1 determinately survives, say, a quarter of the operations, and determinately doesn't survive the others. The semantic candidate for 'same person' being employed when assessing the spectrum would be (determinately) satisfied in only a quarter of the cases, and (determinately) not satisfied in the rest. Now, in his discussion of the Combined Spectrum, Parfit rejects sharp borderlines. But we can

see now that if (a) is false and (b) is true there *will* be a sharp borderlines – but they won't be *substantive* borderlines. In the present example, there would be a borderline a quarter of the way across the spectrum. And it is easy to see that there would be many other such borderlines at various locations on the spectrum, each according to a different but equally legitimate semantic candidate. These borderlines would not reflect 'deep' differences between identity and non-identity; they would be what we might call *nonsubstantive borderlines.*

Let's now think about what (ii) would look like. Again, we are supposing that (a) is false, and so there will be nonsubstantively true answers to identity questions. But we are also supposing that (b) is false, meaning that some of these answers will say that it is indeterminate whether p1 survives. This might sound a little odd at first, but it makes sense. With (a) being false, there will be cases where it is an empty question whether p1 survives. In these cases, we already know everything there is to know. That being so, we could *describe* an operation as one in which identity (determinately) holds, or we could describe it as one in which identity (determinately) doesn't hold. But if it *really is* an empty question, and we already know everything there is to know, then we could equally describe an operation as one in which identity *neither does nor doesn't hold*. We could employ a semantic candidate for 'same person' that resulted in borderline cases of identity where no determinate answer could be given. We could, in other words, employ a *vague* semantic candidate.

For this reason, it is a mistake to claim – as (i) does – that the answers to nonsubstantive questions *cannot* be indeterminate. Simply: if the question is nonsubstantive, it cannot be substantively false that identity is indeterminate. If we already know everything there is to know about an operation, then saying that identity is indeterminate is just another way of *describing* the underlying fact. Disagreements between proponents of these views would, again, be merely verbal; different answers will be true relative to the semantic candidate they employ – which could be either vague or non-vague.[11] So if (a) is false, (b) is also false. We may now discount (i).

[11] This answers a separate concern one might have about (ii). (ii) claims that (a) and (b) are false. The falsity of (a) and (b) means that some questions will be nonsubstantive, and some will have indeterminate answers. This leaves the logical possibility of mutual exclusivity – i.e. that all nonsubstantive questions will have determinate answers, and all indeterminate answers will be answers to substantive questions, and no nonsubstantive questions have indeterminate answers. I do not consider this possibility for the reasons above: if a question is nonsubstantive, it can have indeterminate answers.

There are broadly speaking three sorts of view about how vagueness arises: linguistic, epistemic, and ontic. There is a case to be made for any one of them being the source of a semantic candidate's vagueness. I do not want to choose between them, but will outline what proponents of the different views could say.[12]

According to the linguistic theory of vagueness, vagueness results from imprecisions in our language. It regards vagueness as 'semantic indecision' (Lewis 1986: 212): the reason there are borderline heaps is not because heaps are vague entities. Rather, there are many precise entities, but 'heap' isn't precise in which of those it picks out. A term is vague because of the way we use it; we don't have in mind the exact referents. We could, therefore, employ a semantic candidate for 'same person' that was imprecise in this way, and hold that this semantic candidate was no better or worse than any other.

The epistemic theory also denies that there are vague objects. But it also denies that our words are imprecise. Details differ between accounts, but the general idea is that vagueness is the result of ignorance. Words like 'heap' have precise referents, it's just that we don't *know* what those precise referents are (e.g. Williamson 1992; Sorensen 2001).[13] We could therefore hold that one of the possible semantic candidates was vague in this way. Although there are many semantic candidates that we can specify the referents of, each of which enables us to give a determinate answer to the question of identity, there are some that we can't specify. For these candidates, we simply don't know which entities they pick out. And so we cannot give determinate answers to the question of identity.

According to the ontic theory, vague terms refer to vague entities. It locates vagueness 'in the world'.[14] Lewis's example is the outback: proponents of ontic vagueness think that the reason we can't say where the outback begins is because 'there's this thing, the outback, with imprecise borders' (Lewis 1986: 212). Similarly, the reason we can't say when a collection of sand grains becomes a heap is because heaps themselves are fuzzy things. So while we can employ semantic candidates for 'same person' which, through

[12] The task of choosing between the different accounts would be a project in itself. Moreover, it is not even clear that we need to choose between the three types listed. The list might not be exhaustive for one thing, and the different types of vagueness might not be mutually exclusive (Williams 2008).

[13] The epistemicists are yet to provide an account of the connection between meaning and use. But, Williamson argues, '[s]ince no one knows what such an account would look like, the epistemic view of vagueness should not be singled out for its failure to provide one' (1992: 157).

[14] Traditionally, ontic vagueness was largely disregarded, sometimes on the grounds of unintelligibility (see e.g. Dummett 1975; Lewis 1986; Russell 1923).

stipulation, have precise referents, we can also, if the question is empty, employ semantic candidates which don't have precise referents. We could – if we were willing to entertain the notion of ontically vague entities – stipulate that some of the possible semantic candidates pick out ontically vague entities. And relative to these candidates, it would be nonsubstantively true that identity neither does nor doesn't hold - provided, that is, none of the ontically vague entities to which the semantic candidates refer are in any way metaphysically privileged (because then the question would be substantive).

According to (ii), it will be nonsubstantively true that cases on the spectrum involve indeterminate identity. Assuming that indeterminacy is the result of vagueness, the source of this vagueness might be semantic, epistemic, or ontic; there is, at least, a case to be made for each of them.

We have discussed how, if questions of identity can be nonsubstantive, the spectrum would be populated with nonsubstantive sharp borderlines. Well, we can now see that the spectrum will also, given (ii), be populated with nonsubstantive borderline *zones*: regions spanning many cases on the spectrum for which it is indeterminate whether p1 survives. These borderline zones will reflect nonsubstantively true 'neither yes nor no' answers – answers which are true *relative to the vague semantic candidate they employ*.

In rejecting (a), (ii) allows that questions of identity can be nonsubstantive, or, as Parfit calls them, empty. So far so good for Parfit. But what about (iii)? (iii) accepts (a), so denies that questions are nonsubstantive. This *prima face* looks like a problem for Parfit's conclusion that identity questions can be empty. However, it also rejects (b), and so says that identity questions can have indeterminate answers. When we spell out what this means, it emerges that Parfit has already given reason to reject (iii), and so his conclusion is safe.

If (a) is true and (b) is false, it will be substantively true that identity neither does nor does not hold for (at least) some cases on the spectrum. For this to be substantively true, there must be a semantic candidate that picks out a metaphysically privileged entity in the world for which identity neither does nor doesn't hold in those cases. For the Reductionist, there is no such metaphysically privileged entity. Parfit made the same point about in his objection to a substantive borderline. He asked the Reductionist: what could make it true – *substantively* true – that in one case we had an instance of identity, and in the next an instance of non-identity? The Reductionist had no reply. Only the Non-Reductionist did.

The question is now: *what could make it substantively true* that in some cases we have an instance of *indeterminate* identity? The primitivist can dodge this version of the truthmaker problem in much the same way as before. But the Cartesian is free to hold that Egos are ontically vague entities. This would supply them with the privileged semantic candidate they need for indeterminate answers to be substantively true. But for the Reductionist, there is nothing in the world that could make it substantively true that identity is indeterminate. As such, (iii) is an option only for Non-Reductionists. But as we have already rejected Non-Reductionism, (iii) is out of the picture. For the Reductionist, the only live option is therefore (ii).

We can now see that Parfit's assessment of the Combined Spectrum was somewhat superficial, and potentially misleading as a result. Three of his claims are in need of clarification. Firstly: the claim that there can be no sharp borderline. There *will* be sharp borderlines – many of them – but they will be what we have been calling *nonsubstantive* borderlines. Parfit is correct, however, if we read him as claiming that there cannot be a sharp *substantive* borderline.

Secondly: the claim that identity can be indeterminate. There are three ways in which indeterminacy has cropped up in our discussion of the Combined Spectrum. Firstly, we saw that it cannot be *substantively* true that identity is indeterminate, since that would require a semantic candidate to pick out some metaphysically privileged vague entity – which only the Non-Reductionist can posit. But there are two ways in which indeterminacy *does* feature on the spectrum, but which Parfit fails to distinguish between. Since there is no substantive borderline, there will be a region where there are only *nonsubstantive* truths about identity. And so, in this region, questions of identity will be indeterminate in the sense that it will be *neither substantively true nor substantively false that identity holds*. This, I take it, is why Parfit says that in some cases there is no 'deep' difference between identity and non-identity. Questions of identity will also be indeterminate in this region in the sense that it will be nonsubstantively true that identity neither does nor doesn't hold – i.e. there will be nonsubstantive borderline zones corresponding to (equally legitimate) vague semantic candidates. Because it is an empty question whether identity holds, we can legitimately *describe* identity as *neither holding nor not holding.*

Thirdly: the claim that when a question of identity has an indeterminate answer, that question is *empty*. This claim can be assessed with respect to the three types of

indeterminacy listed above. Firstly, if it were substantively true that identity is indeterminate, this would be because there is some metaphysically privileged vague entity of the Non-Reductionist sort. If this were the case, it would most certainly *not* be the case that the question was empty. There would be a 'deep', non-verbal truth – a truth about the fuzziness of a vague entity 'in the world'. The 'deep' fact of the matter would be that *it is neither true nor false that the entity's identity holds*. As such, you couldn't describe the case either way; you *have* to describe it as one of indeterminate identity. The identity question is not therefore empty if it is substantively true that identity is indeterminate. Since we are rejecting Non-Reductionism, this needn't worry us.

Secondly, there will be cases in which it is neither substantively true nor false that identity holds. In this borderline zone, we saw, there will be only *nonsubstantive* truths. This is why Parfit claims there will be no deep difference between identity and non-identity. There are only *nonsubstantive* truths in this region – answers to identity questions that are true relative to the non-privileged semantic candidates they employ. This is what it is for the question to be empty. Different answers are just different ways of describing the underlying facts.

Finally, within this region of nonsubstantive truths, there will be nonsubstantive borderlines – sharp borderlines corresponding to determinate semantic candidates – and there will be nonsubstantive borderline *zones* – regions corresponding to *vague* semantic candidates. In these nonsubstantive borderline zones it will be nonsubstantively true that identity is indeterminate. Given that it is nonsubstantively true that identity is indeterminate, the question will be empty. But it is not this indeterminacy *per se* that entitles us to conclude that the question is empty. We have already concluded that questions of identity within the region of nonsubstantive truths are empty. And it is *because* they are empty that we may *describe* them as involving indeterminate identity.

2.4 Extending Parfit's claim

Now that we have a better understanding of how the Combined Spectrum argument works, we are in a position to see its shortcomings. In this section, I will argue that Parfit's claim that questions of personal identity are empty in some cases on the spectrum is unjustified. Instead, we should conclude that *all* questions of personal identity are empty. As we have seen, Parfit advocates what Curzer (1991) calls the 'three stage view':

> The resulting person would be me in the first few cases. In the last case he would not be me. In many of the intervening cases, neither answer would be true. (Parfit 1987: 232-233)

However, I have argued that Parfit may only conclude that questions of identity can be empty from the claim that identity can be indeterminate when we interpret that indeterminacy claim as:

> It is neither substantively true nor substantively false that identity holds

We may therefore modify Parfit's quote in such a way as to make this indeterminacy claim explicit. Thus, I propose we understand the three stage view as:

The three stage view:	In the first case, it would be substantively true that identity holds.
	In the last case, it would be substantively false that identity holds.
	In many intervening cases, it would be neither substantively true nor substantively false that identity holds.

I will assume this clarified version of the three stage view in what follows. At the near end of the spectrum is a zone of substantive identity, at the far end is a zone of substantive non-identity, and between these is a borderline zone of nonsubstantivity.

There are a few problems with the three-stage view. Firstly there is the difficulty of how to deal with the boundaries of the nonsubstantive zone. As Curzer (1991) sees it, there are two options. Either the central nonsubstantive zone has sharp borderlines, or it is fuzzy. If it has sharp borderlines then clearly it runs into the same trouble as the view Parfit is arguing against. The difference between p1 persisting and it being an empty question whether p1 persists would consist in some trivial difference, and, again, the borderline would be completely unobservable. I agree with Curzer that the nonsubstantive zone cannot therefore have sharp borderlines. However, according to Curzer, the fuzzy three-stage view is no better off:

> Suppose the Empty Area [i.e. the nonsubstantive zone] has fuzzy borderlines. What does it mean to say that F is a point in a fuzzy area? F is not in the Empty area, so it is not true that […] it would convey no additional information to say that person A and person F are the same or different. In other words, it *would* convey additional information to say that person A and person F are the same or different. But this implies that F is in either the Same area or the Different area. Thus, there is no Fuzzy area. The empty area has sharp borderlines. (Curzer 1991: 22-23)

I do not think we should be persuaded by Curzer's argument here. We are imagining a case on the spectrum that lies in the borderline (fuzzy) zone between the central nonsubstantive zone and one of the substantive zones at either end of the spectrum. Now, there are three different views about what attitude one should have towards such borderline cases. Consider the proposition:

P: 'the case in question lies within the nonsubstantive zone'

Some people would say that P is neither true nor false; some would say P is both true and false; and other would say that P is *partially* true. Following Williams (2016), we can call these theorists *gappers, glutters, and scalars* respectively. Gappers should fully reject P, glutters should fully accept P, and scalars should accept P to whatever degree they think it is true (*ibid.*). Curzer says that the case does *not* lie within the nonsubstantive zone ('F is not in the Empty area'), and in doing so appears to be fully rejecting P. We should therefore take Curzer to be a gapper, and I assume that approach in what follows.

To see where Curzer goes wrong, consider first what we should say if the spectrum case in question *was* in the central nonsubstantive area. If it was in the nonsubstantive area, it would be an empty question whether or not identity was preserved. Answers to empty questions do not tell us about the way the world is; they merely tell us about how the person who gives the answer *describes* the world – or, as Curzer puts it, answers to empty questions 'convey no additional information'. If it is *not* true that F is in the nonsubstantive area, then it will not be true that the identity question is empty; and so it will not be true that answers to that question would convey no additional information. Curzer is therefore quite entitled to say:

(iv) it is not true that answers to the identity question would convey no additional information

The problem, though, is that he thinks (iv) entails:

(v) answers to the identity question *would* convey additional information

For (v) to follow from (iv), however, it must be the case that:

(vi) it is *false* that answers to the identity question would convey no additional information.

But we cannot accept (vi), since we are supposing that it is *neither true nor false* that the case in question sits within the nonsubstantive zone – that it is neither true nor false that the question is empty. If it is neither true nor false that the question is empty, then:

(vii) it is *neither true nor false* that answers to the identity question would convey no additional information.

(v) does not follow from (vii), and so Curzer's objection fails. There is however another obvious stumbling block in the way of the three-stage view: the problem of infinite regress. To avoid sharp borderlines, we will need to say that the edge of each borderline zone is itself a borderline zone *ad infinitum.* A staunch defender of the three-area view might be willing to accept this regress. But the real problem is: how wide are these borderline zones? Well clearly they must decrease in size, by which I mean they must span a decreasing number of cases on the spectrum. Let's suppose the central borderline zone (b1) that Parfit posits in the three-stage view spans n number of cases. A borderline zone of b1, b2 will then span some fraction of n, and b3 will span some fraction (presumably the same fraction, but this is not vital) of that. Also, we may imagine the spectrum to be dense, such that there are an infinite number of cases (it could be taken down to the subatomic level and so on). So the width of the borderline zones will look something like:

b1……..... n

b2 ……..……. $\frac{n}{x}$

b3 ……..…. $\frac{n}{x^2}$

$b\infty$ ……. $\frac{n}{x^{\infty}}$

Thus, the width of the borderline zone adjacent to either substantive zone will be approaching zero, but the cases either side of that borderline will be only trivially different. Hence, the fuzzy three-stage view will be prone to the same criticisms (or at least, very close relatives of the criticisms) facing the sharp three-stage view. Parfit acknowledge the difficulties of the three-stage view, if only to admit defeat:

> There is a problem here, that of how to handle the borderline cases of borderline cases. I do not know how to solve this problem. But I am sure that it cannot show that there are *no* borderline cases. It cannot show that, to all questions, in all conceivable cases, there must be determinate answers. (Parfit 1993: 31-2)

I agree. Certainly, there is no need to concede that questions of identity are always non-empty. But Parfit is ignoring another possibility here. He is ignoring the possibility that *all* questions of identity along the spectrum could be empty. This is what Curzer calls the 'one stage view' (1991). It is curious that Parfit doesn't argue for the one stage view, since it can be seen to follow from Reductionism (as noted by Curzer 1991).

By Reductionism, there is nothing more to a person's existence than the empirical facts. As such, if we knew those facts we would know everything there is to know about that person. Consider the question:

> If changes [x, y, z] are applied to p1's empirical properties, does p1 persist?

This question will always be empty. Because there is nothing more to a person's existence than the empirical facts, answering the question will convey no additional information. As such, *any* case on the spectrum will be indeterminate in the empty-question sense. Parfit restricts this reasoning to the middle cases of the spectrum, but it applies equally to all cases along the spectrum. To employ Parfit's arguments against him, his claim that the first cases are cases of determinate identity and the last are cases of determinate non-identity is merely a reflection of how he describes those cases.

Curzer objects that the one-stage view doesn't correspond to ordinary language. This is a bit too quick. That all questions along the spectrum are empty doesn't mean you can't say

that persons persist through change. Ordinary language is perfectly admissible. It's just that you *could* say they *don't* persist. Neither answer reflects a 'deep' truth about personal persistence, because there aren't any truths over and above the empirical facts. If the objection is rather that the one-stage view doesn't reflect our intuitions about the metaphysical significance of personal persistence, then this is no objection at all. The whole point of the Combined Spectrum is to show that these intuitions are faulty. If one doesn't want to accept the one-stage view, their only option is to embrace Non-Reductionism.

The conclusion is that, for *any* amount of change, whether or not a person persists is an empty question. We say that persons persist through change only as a matter of linguistic convention; any change could be instead described as involving distinct persons. It is fair to assume that, for any amount of time, a person's empirical features will change in some way, no matter how trivial. Hence, it is merely linguistic convention that persons persist *per se.* What we naturally describe as a single persisting person could instead be described as a series of distinct momentary persons. We can say that persons persist, but this is only nonsubstantively true; it is only true relative to the semantic candidates we employ.

At this point there may be concerns that what is true of persons is true also of any entity we can give an empirically informativist (Reductionist) account of. For example, the existence of a river just consists in lower-level empirical facts, and so is open to a spectrum-type argument. Accordingly, we should regard the diachronic identity of rivers as a merely verbal matter. The same holds with nations, and trees, and so on. I am not worried by this conclusion. As I see it, these are the *easy* cases. The ancient question of whether the same river can be stepped in twice is an empty question, and this is not something that normally troubles us. Of course, there are practical benefits that come with *saying* it is the same river. Similarly, there are benefits that come with reidentifying nations, and trees. But our conclusion – that the diachronic identity of such things is merely verbal – should surely be *less* troubling here than in the case of personal identity. Unlike with rivers, our intuition that there is a deep truth about personal persistence is given to us by our first-person perspective. And so if this intuition is shown to be mistaken, the hard work is done.

2.5 Experiential identity and empty questions

Given the truth of Reductionism, it is an empty question whether persons persist. I now want to use this claim to argue that the same can be said of experiences. Our starting point will be the common sense idea that experiences are individuated by subjects, such that two distinct persons cannot be subjects of one and the same experience.[15] The thought is put by William James as follows.

> Neither contemporaneity, nor proximity in space, nor similarity of quality and content are able to fuse thoughts together which are sundered by this barrier of belonging to different personal minds. The breaches between such thoughts are the most absolute breaches in nature. (James 1952: 147)

Even if you and I are having an experience *of* the same thing – maybe we are both watching the same film at the same time etc. – we certainly wouldn't want to say that there is only one experience occurring, but one that has two subjects. There would be two distinct experiential *events*, but which both had the same *content*. Pursuing this line of thought, Barry Dainton imagines a replica is made of you, complete with an exactly similar sequence of experiences:

> Consider the following sequences of phases: S1 = [m1-m2-m3 … -m20], S2 = [**m1**-**m2**-**m3** … **-m20]**, and S3 = [m1-**m2**-m3-**m4**-m5-**m6**-m7 … **-m20**]. The first two sequences consist (respectively) of a portion of your experience and a portion of your replica's. The third sequence consists of an alternation of phases from both streams of consciousness, yours and your replica's. Clearly, S3 is not a real stream of consciousness at all; it is a *fictional amalgam* of components from two real streams, S1 and S2. (Dainton 2006: 130, emphasis added)

Intuitively, I persist through time, and my stream of consciousness accompanies me. It is entirely natural to suppose that my experiences are temporally extended in the same way that I am, and that this is a 'deep' truth – just as that of my persistence is. However, since the intuition about personal identity can be challenged, so can the intuition about experiential identity.

What we usually describe as a single persisting person could equally be described as a series of distinct momentary persons. But if it is impossible for two distinct persons to be

[15] For the purposes of this discussion, 'person' and 'subject' can be used interchangeably: the experiences of two distinct persons/subjects cannot be unified into a genuine whole.

subjects of one and the same experience, then, were we to adopt this alternative way of talking, we would have to deny that experiences are ever temporally extended. On the assumption that the experiences of distinct subjects can't be unified, we would have to say that experiences are only ever momentary states – each of which belongs a distinct momentary subject.[16] We would have to say that the extended experiences we normally talk about are merely 'fictional amalgams'.

So whilst we can *describe* persons as persisting, and as having temporally extended experiences, we could equally describe them as momentary, and as having *momentary* experiences. Since it is an empty question whether persons persist, it is an empty question whether experiences have temporal extension.

2.6 Experiences without subjects

Call the common sense idea that experiences are temporally extended in a 'deep' non-verbal sense 'extensionalism'.[17] The extensionalist might be prepared to admit that, when we look too closely at the *subjects* of extended experiences, common sense starts to look rather odd. But they might argue that this is to start off on the wrong foot. Sure, when we think about how *subjects* fit into the picture, it seems we will be forced to admit that it is an empty question whether experiences are temporally extended. But the extensionalist might want to insist that it is only the concept of 'subject' (or 'person') that is to blame – that it is somehow defective, and that this defect is infecting the way we think about experience. They might want to insist that we should amputate the idea of the subject from the way we think about experience. They might want to do away with the concept of 'subject' altogether, and thus avoid the Parfittian problem.

[16] A reminder: I am using the terms 'experiential *state'* and 'experiential *event'* idiosyncratically. When I use the term 'experiential state', I simply mean an experience that has no temporal extension. This is *not* a claim about its content. An experiential state, as I intend it, might represent a duration or an instant. An experiential event, as I use the term, is simply a temporally extended experience. So what Dainton calls a 'stream of consciousness' would be an experiential event in my books.

[17] The name 'extensionalist' is standardly used to refer to those who posit extended experiences in order to account for our perception of temporal phenomena (such as Dainton 2006; Phillips 2011a). Used in this standard way, extensionalists will hold that experiences are temporally extended in a 'deep', non-verbal sense – and so would accept the title as I am using it. However, I am using the title here in such a way that remains silent on the debate about our perception of temporal phenomena. For now, I simply use 'extensionalist' to refer to those who hold the common sense view about the temporal extension of experiences.

This would be a reasonable response, I think. In arguing against the extensionalist, we adopted the Parfittian stance that persons (subjects) lack the sort of metaphysical status we intuitively attribute to them. But Parfit (1999) himself developed this line of thought further, going on to argue that a conceptual scheme that lacked the concept of 'subject' would be no worse off than ours, metaphysically speaking. *Our* concept of 'experience' demands the existence of subjects, insofar as experiences are the sorts of things that are *had.* Since experiences are *had,* there must be subjects *having* them. But whilst it is true that there are such things as subjects, this is true only because of the way we talk about *experiences*. We can, however, imagine an *impersonal* conceptual scheme. This impersonal conceptual scheme would feature an analogue of 'experience' – 'experience*' – that made no such demand. There would be such things as experiences*, but we wouldn't think about *ourselves* as the subjects of those experiences. Statements like 'I am happy', would be replaced by statements like 'this is an expression of an experience* of happiness'. The concept of 'subject' is, on this view, metaphysically superfluous.

The extensionalist may therefore want to fully embrace this impersonal conceptual scheme in an attempt to overcome the problem facing their view. Indeed, if they were to agree, they would presumably want to claim that an impersonal conceptual scheme would be in some ways *superior* to ours. When it comes to thinking about temporal experience, the concept of 'person' just gets in the way. Like the appendix, it is obsolete and liable to cause problems; we would be better off with it removed. This would be good news for the extensionalist if it can be made to work. The extensionalist has found themselves cornered by an argument derived from a Parfittian line of thought. But it now looks like this Parfittian line of thought might also, when followed a bit further, provide the means of escape.

The remainder of this chapter will examine the prospects for extensionalism when persons are taken out of the picture. I'm not going to *talk* in impersonal terms – it's clunky and unnatural, and so makes things hard to follow. But what follows could be translated into an impersonal conceptual scheme.

2.7 What experiential identity consists in

The extensionalist is by no means out of the woods. Even if they want to champion the impersonal conceptual scheme, they will still owe us an account of what unites successive

momentary experiential states into a genuinely extended experiential event. Dainton (2006) calls the relation in virtue of which experiences are unified over time of 'co-consciousness'. In what, then, does the relation of co-consciousness consist?[18]

It should, I hope, be clear that any sort of psychological and/or physical continuity account will be susceptible to the Combined Spectrum argument. For example, one might be tempted to suppose that the physical continuity of the brain is necessary and sufficient for experiential identity, such that, even if the content of two experiences was radically different, they would still be united in virtue of the fact that they were 'generated' (however you want to take that) by the same brain – where 'same brain' would presumably be unpacked in terms of physical continuity. Or, one might be tempted to suppose that co-consciousness consists in psychological continuity, such that, if a single brain suddenly suffered severe trauma and the memories etc. were lost, there would be a break in the stream of consciousness. We would be talking about two distinct extended experiential events. Each of these sorts of view has its intuitive appeal, I think. But clearly we could run the Combined Spectrum argument (I trust we don't need to go through the reasoning again), replacing the question of personal identity with the question of experiential identity, and end up at the same place: questions of experiential identity are empty.

There is, however, one view that might be thought to offer a means of avoiding the Combined Spectrum argument: the phenomenological account of co-consciousness. This is the account proposed by Barry Dainton (2006). We can introduce the idea by first considering Russell's (1921) claim that we cannot, on the basis of present experience alone rule out the hypothesis that the world sprang into existence five minutes ago. So long as we have the appropriate memories, we would be none the wiser. We can modify this suggestion. Instead of five minutes ago, let's make it a bit more extreme. Let's say a moment ago. We perceive the world in its current state – state Q – and have the false memory of the world in the preceding state, P. The world never was in state P;

[18] I assumed in the previous section that the experiences of two distinct subjects cannot be unified into a genuine whole. Now, there may be those that want to deny this assumption. There are theorists who think that distinct subjects' consciousnesses *can* be combined, such as those who espouse a version of panpsychism (e.g. Luke Roelofs 2016) or cosmopsychism (e.g. Philip Goff 2017). Even if we were to entertain the possibility that different subjects' experiences could be unified, however, we would still need an explanation of how experiences combine in the particular way required here – i.e. between subjects diachronically related by the particular empirical relation(s) deemed necessary. Further, we would still need an account of what the relation of co-consciousness consists in – the difficulties which are explored presently.

nevertheless, provided we have the appropriate (false) memories, we wouldn't be able to tell.[19] The world could have sprung into existence *just now*.

Dainton disagrees. He thinks we *would* be able to tell if the world had suddenly come into existence. Or, at least, he thinks there would be phenomenological evidence of it (whether we would conclude from this evidence that the world had just now sprung into existence is another matter). There would be evidence because an essential phenomenological ingredient would be missing – an ingredient that isn't found in memory (fake or otherwise). He tries to muster support for this idea by imagining what it would be like to not have any memories *whatsoever*.

> Suppose you have been given a drug which over the course of a day gradually and sequentially destroys your experiential memory, without impairing your mental functioning in any other way […] Although you are now stranded in the immediate sensory present, you hear sounds, you have bodily sensations, you see movement, you feel strangely bewildered. The phenomeno-temporal character of immediate sensory experience is much as it ever was. True, it is impossible to imagine with any clarity what it would be like to be in this condition; but it is easy to believe that one could continue to be conscious, and continue to be directly aware of change. (Dainton 2006: 124)

Dainton holds the common sense view that experiences are temporally extended in a metaphysically substantive sense. He is therefore an 'extensionalist' in the weak sense that I have been employing. But he is also an extensionalist in the stronger sense – he holds that it is only in virtue of having extended experiences that we perceive change. In the passage above he is claiming that we could perceive change without memory. Since he holds that change is represented in extended experiential events, he holds that memories aren't necessary for extended experiences; co-consciousness can hold in the absence of memory. Rather, he thinks that what the thought experiment shows is that a different type of continuity grounds co-consciousness: *phenomenal* continuity. Phenomenal continuity is not, Dainton insists, the same as psychological continuity. Psychological continuity is unpacked in terms of memories, beliefs, desires, and so on.

[19] Dainton's version of the thought experiment is one in which, instead of springing into existence a moment ago, the world emerges from a 'thousand-year total freeze, a period during which all mental and physical processes were completely halted' (2006: 131). His conclusion transposes to the example I imagine.

But phenomenal continuity 'cannot be reductively analysed (at least in any remotely obvious way) in terms of beliefs and memories' (Dainton & Bayne 2005: 554).

Consider again Dainton's examples of two genuine streams of consciousness, S1 and S2, and the 'fictional amalgam', S3. The reason S3 isn't a genuinely extended experiential event – even though you and your replica have the same memories, beliefs and desires – is because the experiential states it comprises aren't phenomenally continuous:

> If we know anything about consciousness, we know that there is a fact of the matter as to which experiences [i.e. momentary experiential states] belong to which streams [extended experiential events], and we know this in virtue of the purely phenomenal interconnections between co-streamal experiences. (Dainton 2006: 130)

The relation of phenomenal continuity is one that is *felt*. It is felt when an experience 'merges seamlessly with the next' (Dainton & Bayne 2005: 553). The mistake we make in the Russell-type thought experiment, Dainton thinks, is to suppose that our experience of the world in state Q would have the same phenomenal properties whether or not that experience was preceded by a genuine experience of P. Because the previous experience would be absent, the relational property would not obtain. So if the world were to have suddenly popped into existence in state Q, our experience of Q would not seem to flow from our (falsely remembered) experience of P. Although you would seem to *remember* the preceding experience, your current experience would not seem to have 'merged seamlessly' from it.[20] This phenomenal unfamiliarity would be a 'tell-tale' sign that the world had just popped into existence. This is, of course, speculation on Dainton's behalf. But let's give him the benefit of the doubt here, and see how far we can run with the idea.

There are two related objections I want to raise. Firstly, that the phenomenal continuity account relies on some pretty dubious empirical assertions, and secondly, that the view amounts to a version of Non-Reductionism about experiential identity, and should be rejected as such.

[20] I think we might be able to get a feel for what Dainton is suggesting here if we try to imagine a sort of inverse déjà vu. Where déjà vu is a sense of phenomenal familiarity ('it somehow feels like I've had this experience before), you would have a sense of phenomenal *unfamiliarity* ('this somehow doesn't feel like the experience I was having before').

2.8 Rejecting the phenomenal continuity account

That the phenomenal continuity account of co-consciousness amounts to Non-Reductionism about experiential identity can be brought out by another thought experiment: the 'streamal diverter' (Dainton & Bayne 2005). The streamal diverter is a piece of apparatus that that can 'direct the flow of consciousness from one brain to another in an instant' (Dainton & Bayne 2005: 554). Since phenomenological continuity is – we are assuming – distinct from physical and psychological continuity, we can imagine that the memories, beliefs and desires 'housed' in each brain remain in place, unaffected after the swap. In other words, we are imagining a case where phenomenological continuity cuts across physical and psychological continuity. Consider the two brains and the associated sequences of experiences:

Brain 1, generating the sequence of experiences e1, e2, e3, e4

Brain 2, generating the sequence of experiences e1*, e2*, e3*, e4*

In everyday life, we imagine that each brain has its own stream of consciousness. The stream of consciousness associated with Brain 1 would, we suppose, comprise experiences e1 – e4, and the stream of consciousness associated with Brain 2 would comprise experiences e1* – e4*. We can imagine that Brain 1 is receiving sensory input from – and so is giving rise to experiences of – an opera hall in which the Marriage of Figaro is being performed; whereas Brain 2 is receiving sensory input from – and so is giving rise to experiences of – a painful root canal extraction.

Suppose that a streamal diverter is used to swap the streams of consciousness between brains. There would then be two streams of consciousness that looked like:

Stream 1: e1, e2, e3*, e4*

Stream 2: e1*, e2*, e3, e4

Dainton and Bayne think that personal identity goes with experiential identity, and for the purposes of discussing this thought experiment it will help to talk in these terms (although what is said could be equally described in impersonal terms).

The prospect that 'streams of consciousness' could be swapped like this should strike us as a peculiar suggestion. After all, we are supposing that *psychological* continuity would remain with the original brain, meaning that all memories of the experiences leading up to the streamal diversion would be lost. Suppose 'your' stream of consciousness is Stream 1. After having your stream of consciousness diverted from the opera house to the dentist's chair, it would seem that you had been at the dentist's all along. You would remember sitting in the waiting room filled with dread; being called in by the nurse; lying back in the chair and so on. All memory traces of the opera would be gone.

To accept this story we have to square two other of Dainton's empirical assertions:

Claim 1: Even without any memory of what happened in the preceding moments we would nevertheless still feel the 'flow' of phenomenal continuity (as in the thought experiment in which a drug destroys your memory)

Claim 2: If we were to have false memories of what happened in the preceding moments we would feel the 'jolt' of phenomenal discontinuity (as in the Russell-type thought experiment).

We have been told that the successive experiences within a stream of consciousness are unified by virtue of being phenomenally continuous – when one experience 'merges seamlessly with the next'. But it is obscure to say the least how the streamal swap would feel in seamless in the required way.

Since e3* would be accompanied by a memory impression of e2*, we would believe that the preceding experience was e2*. But, since e2* and e3* are not part of the same stream – they are not co-conscious – *ex hypothesi*, e3* would not seem to have merged seamlessly from e2*. Yet at the same time, because e3* is co-conscious with e2, there *would* be a sense of phenomenal flow – but, since we would have no recollection of e2, we wouldn't be able to say what e3* seemed to flow *from.* So we would feel a strange sense of phenomenal 'dislocation'. This is, of course, pure conjecture. These are empirical claims that lack any sort of supporting evidence, and so should be treated with scepticism. This in itself counts against the view. But let's try to keep an open mind. It will come in useful for what I now want to show: that the view amounts to Non-Reductionism about experiential identity.

Dainton and Bayne's thought experiment involves two brains and two sequences of experiences. But let's now imagine a streamal diverter is used on four brains and their respective sequences of experiences. Also let's imagine that we don't know in advance the destinations of each stream of consciousness. We would have a situation in which four people are reporting to feel a strange sense of phenomenal dislocation. They would report (if we are to believe Dainton and Bayne) that the experience immediately following the activation of the streamal diverter did not seem to 'flow' from the previous one; but that the experience immediately after the activation of the machine nevertheless did seem to flow from some unidentifiable source. If Dainton and Bayne are right, then there will be a fact of the matter about which brains inherited which 'streams of consciousness'.

But now what if we imagine a different machine is made. Instead of transferring streams of consciousness from one brain to another, this machine induces a feeling of phenomenal dislocation of the sort discussed. We again have four test subjects who provide the same phenomenological reports as in the previous experiment. We may ask the question: what is the difference between the two experiments? In both the experiments we know what the physical and psychological facts are. And if we are to count the phenomenological facts separately, we know what these are too. All these empirical facts are the same in both cases.

Dainton would object to this last point. Dainton (2008: 70) regards the phenomenological properties in question as *relational* properties, in the sense that they bind successive experiences together. It is only in virtue of being the subject of *successive* experiences that we have this sense of flow. In order to feel the phenomenal flow from one particular experience to another, we must be the subject of *both* experiences. Phenomenal continuity cannot be experienced in a single momentary experiential state. Hence, the empirical facts won't be the same in both cases.

This seems like a reasonable enough claim, but becomes less so upon inspection. Consider two temporally adjacent experiences (or experience-parts), e1 ending at t1 and e2 starting a moment later at t2. If e1 and e2 are phenomenally continuous, we are to suppose that e1 will seem to merge seamlessly into e2. But how is this meant to work? At t1, e2 is yet to happen, and so it cannot seem that e1 merges into e2 *before* t2. The transition from e1 to e2 can only seem smooth once e2 is being experienced – *viz.* at t2. So, speaking precisely, e1 cannot seem to merge *into* e2. It's rather that e2 will seem to *have* merged *from* e1. The sense of phenomenal flow cannot therefore be relational in Dainton's sense.

It must rather be what we might call a *retroactive* property. The merging of one experience into another can only seem smooth during the second experience, *after* the first experience is over.[21] On the assumption that phenomenology at a time supervenes on neural state at a time – which Dainton (2008: 324) appears to accept – the sense of phenomenological dislocation could indeed be induced in a subject.[22]

There will be, therefore, no empirical difference between the two cases. In terms of empirical facts –facts that are discoverable – we must describe the two cases in exactly the same way. Dainton and Bayne want to insist that there *is* a difference between the cases, though. They want to posit *further facts*: facts about which experiences belong to which streams (or, which momentary experiential states belong to which extended experiential events). These are facts for which there can be no evidence, however. In the same way that we have no evidence to believe in the existence of separately existing Cartesian Egos, we have no evidence to believe in the existence of separately existing streams of consciousness. And in the same way that we cannot be justified in holding that the fact of a person's diachronic identity is a *primitive* fact, we cannot be justified in holding that the fact of an experience's identity is a primitive fact. So their view is indefensible, and should be rejected along with Non-Reductionism about persons.

2.9 Summary

Parfit deployed his Combined Spectrum argument to show that it can be an empty question whether persons persist. Imprecisions in the argument were exposed – particularly regarding vagueness – but, after some shoring up, it was shown that the argument justified an even more radical conclusion than Parfit's: it is always an empty question whether a person persists. And, likewise for experiences. Experiences have temporal extension, but only because of the way we talk. It would be no less metaphysically accurate to describe experiences as strictly momentary states. The move to abandon the concept of 'subject' provides no defence against this conclusion, since the Combined Spectrum is also effective against experiential, rather than personal, identity.

[21] It will do no good appealing to intervening experiences to explain the sense of flow, since in Dainton and Bayne's thought experiment the streamal diverter works instantaneously.

[22] This also provides an argument against Dainton's claim that there would be phenomenological evidence of the world popping into existence a moment ago. Since the phenomenological flow of co-consciousness can't be relational in the way he urges, then, given that we were in the appropriate neural state, there would be no phenomenological 'jolt'.

Finally, the phenomenal continuity account of co-consciousness was considered as a potential means of escaping the Combined Spectrum, but was exposed as a Non-Reductionist view of consciousness. Given the truth of Reductionism, we may conclude that persons and experiences have temporal extension only by virtue of linguistic convention.

Chapter 3: Defending Experiential Atomism

I argued in the previous chapter that given the truth of Reductionism it is an empty question whether persons persist through empirical change, and as such, that it is an empty question whether persons persist *per se.* I also argued that because of this, it is an empty question whether or not experiences can be said to have temporal extension. This claim will now be defended against a view that I shall call experiential holism. According to the holistic conception of experience, the most basic building blocks of consciousness are temporally extended. The idea is that we cannot attribute an experience to a subject *at* an instant other than as part of some interval over which we can attribute an experience. Against the holistic conception stands the atomistic conception, according to which the most basic building blocks of experience are instantaneous.[1] A subject's experience over an interval is always determined by, and so can be completely explained in terms of, that subject's experiential states at each instant during that interval.[2]

The difference between the two views is sometimes put in terms of metaphysical priority. The holist thinks that extended experiences are metaphysically prior to instantaneous experiences, whereas the atomist thinks that instantaneous experience are metaphysically prior to extended experiences.[3] If extended experiences have metaphysical priority, then

[1] I regard holism as a rejection of the atomistic conception that what a subject is conscious of over an interval is *always* determined by what they are conscious of at each instant. That is to say, a holist will think that there are cases where what a subject experiences over a duration determines what they experience at an instant. Whether a holist thinks this is *always* the case is another matter.
[2] I refer to any instantaneous mental item as a 'state'; and a temporally extended mental item as an 'event'. I attach nothing else of significance to the terms.
[3] Much debate in the philosophy of temporal experience is between what are sometimes called 'retentional' (as propounded by Husserl 1950, for example), 'extensional' (e.g. Dainton 2013b; Phillips 2011a), and 'cinematic' or 'snapshot' models (e.g. Chuard 2011; see also Prosser 2017). See Kon and Miller (2015) for an overview. The lines of demarcation between these models become blurred on close inspection, however. One such blurring is between the extensional and cinematic models. Extensional models account for our perception of temporal phenomena by positing temporally extended experiences, whilst cinematic models posit a series of distinct momentary experiences which together form an awareness of temporal phenomena over time. The difference between them then appears to turn merely on whether or not we *call* a conjunction of momentary experiences a single unified experience (Hoerl 2017). The question of metaphysical priority can be seen to more decisively cut between the positions.

it won't make sense to talk about any particular instantaneous experiential state as divorced from an extended experiential event:

> What does it mean to say that a duration of experience is metaphysically prior to its sub-parts? It is not to deny that there are facts about instants during our stream of consciousness. It is, however, to insist that such facts are derivative. The most basic facts about our experiential lives are, in the first instance, facts about extended periods of the stream of consciousness. What is true at an instant is true only in virtue of that instant being an instant during a certain period of experience. (Phillips 2011b: 398)

If true, this would undermine the conclusion from Chapter 2. We would only be able to make sense of a subject having an experience *at* a time once we had *already* helped ourselves to the idea of that subject having an experience that extends *through* time. In picking out an experiential state we would already have implicitly picked out the experiential event to which it belongs. If on the other hand it is the instantaneous experiences that have metaphysical priority, we can pick out an experiential state in isolation. So, for the conclusion of Chapter 2 to stick, the atomist conception must be true.

This can be brought more sharply into focus when we realise that the holistic conception of experience amounts to primitivism about experiential identity. Primitivism about personal identity, if you recall from Chapter 1, is the Non-Reductionist view that the fact of a person's identity consists in no other facts. Similarly, primitivism about experiential identity is the view that the fact of an experience's identity consists in no other facts. Consider first the atomist's view. The atomist holds that instantaneous experiential states are metaphysically prior to extended experiential events. They think that instantaneous experiential states are the most fundamental building blocks of experience, and so they will hold that the fact of an extended experiential event's diachronic identity just consists in facts about instantaneous states. They will think that an experiential event extends through time by virtue of comprising states that are related in some particular way(s). Someone who holds that experiential states are fundamental will therefore be an informativist about the diachronic identity of experiential events. They will think that something informative can be said about what experiential identity consists in.

The holist, however, will have to deny that the diachronic identity of experiences consists in facts about states. That would be to treat the states as more fundamental than the events. But if the states that comprise the extended experiential event can't be used in the

explanans, then no relation that holds between those states can be used in the explanans either. So in holding that extended events are fundamental, the holist will have to deny that the fact of an experience's diachronic identity consists in facts about relations between its constituent states.

It was argued in Chapter 1 that primitivism about personal identity (and Non-Reductionism more generally) should be rejected for epistemological reasons. And only once we have rejected Non-Reductionism can we conclude that questions of personal – and experiential – identity are empty. If it turned out that primitivism about experiential identity was true – as it would if the holistic conception was true – then the conclusion from Chapter 2 would be unwarranted. We could not claim that questions of experiential identity were empty.

The holistic conception has enjoyed the recent support of Ian Phillips (2011a) and Matthew Soteriou (2013). Both offer ingenious arguments that purport to show that the atomist in unable to make sense of motion perception; and that only by adopting the holistic conception are we able to avoid the puzzles they face. Their arguments would, if successful, provide a reason to accept primitivism about experiential identity, and to reject the claim that questions of experiential identity are empty.

I will consider their arguments, showing that empirical evidence in cognitive science gives us good reason to reject their proffered solutions; and moreover, that it is in fact the atomist who is best placed to make sense of motion perception. Before I look at Soteriou's and Phillips's arguments, I will defend the holistic conception from a recent objection brought against it by Simon Prosser (2016). Doing so will also help clarify the holist's position.

3.1 An initial defence of the holistic conception

To illustrate the metaphysical priority of extended wholes, holists sometimes appeal to activities such as walking (e.g. Soteriou 2007). Plausibly, it is impossible for a person to walk for only an instant. Nevertheless, provided a person is walking over some interval, it will be true that they are walking at an instant during that interval. And so we can say that the extended act of walking is metaphysically prior to the action-stages it comprises; the extended item, rather than its instantaneous parts, is fundamental. According to the holist, the same is true of experience.

One might wonder whether such analogies establish quite what the holists want them to, however. Simon Prosser (2016) thinks they don't. All they show, he thinks, is that an instantaneous part cannot exist in isolation. An instantaneous walking-stage cannot *be* an instantaneous walking stage unless it is surrounded by other instantaneous walking-stages. But this dependency fails to establish any metaphysical priority:

> Suppose that an A cannot be ϕ unless it is followed by a B. This does not show that the A and the B together form a whole, or that this whole is more fundamental than its proper parts. For example, no one can be the second person to set foot in Antarctica unless there was a first person to set foot in Antarctica, but this does not show that the first and second persons form a whole. (Prosser 2016: 151)

We may wonder, however, whether Prosser's analogy is entirely fair. It would indeed be rather odd to suppose that the first and second *persons* together form a fundamental whole. But the dependency in the Antarctica example surely doesn't concern persons *per se*; it concerns their *ordering*. The point would be that the position 'second' cannot *be* the position 'second' without being preceded by the position 'first'. More generally, a position in an ordering cannot *be* a position in isolation from the other positions. It is *essential* to the existence of a position that there exists an ordering for it to be a position *in.* Clearly, the same cannot be said for the persons who occupy those positions; a person will exist whether or not they occupy a position within an ordering. So Prosser is quite right to say that the first and second persons are not together more fundamental than they are individually. But the holist would presumable want to reply that what his example shows is that the second person depends on the first person *for their position in the ordering.* So the dependency in Prosser's example might indeed be thought to establish a metaphysical priority – just not the one he attacks.

To reinforce the point, consider again the example of walking. A person-stage can only be said to be walking if it is temporally adjacent to other person-stages that can also be said to be walking. But it would be wrong to conclude from this alone that the *person-stages* are together more fundamental than they are individually. It does not follow simply from the fact that walking is a necessarily temporally extended *action* that persons are

fundamentally temporally extended *beings*.[4] And this is because the dependency in question does not concern person-stages; it concerns their actions (or action-stages).

Perhaps one might still want to resist this claim. Perhaps there is a case to be made for denying that the dependencies in the walking and ordering examples establish any sort of metaphysical priority. Be that as it may, it is not at all *obvious* that they don't – not in the way that Prosser's example was intended to show. So in the absence of any other argument, we have as it stands no good reason to reject the idea that extended wholes *can* be metaphysically prior to their parts. It is of course another question whether this story can be told about experiences. We may now consider the case in favour of the holistic conception of experience: Phillips's and Soteriou's arguments. Soteriou's argument can be dealt with more quickly, so that is where we shall start.

3.2 Soteriou's argument

Soteriou argues that there are cases where a subject's mental state *at* a time may be determined by their mental state *over* time (2013: 106). When reading this claim it should be noted that Soteriou is using 'state' in a different way from the way in which I have been using it. I have been using 'state' idiosyncratically, to refer to things *at* a time, or 'time-slices'. Soteriou is using 'state' in a way that will be more familiar. According to Soteriou's usage, a state can be temporally extended or unextended; but a key feature of states is that they are 'non-dynamic situations' (Rothstein 2004: 6) that 'do not unfold over time' (Soteriou 2013: 102). An example is happiness. A subject can be in a state of happiness at an instant and over time, but this happiness is not processual; it does not *unfold* over time but rather *obtains* throughout an interval.

What Soteriou would describe as an extended state of happiness would therefore, according to my definition, be described as a happiness *event*. So when Soteriou claims that the mental state a subject is in *at* an instant can be determined by the state they are in *over* an interval, he appears to be expressing the holistic conception. He appears to be rejecting the atomistic conception that that a subject's mental properties over an interval

[4] The inference *could* be made, but only with additional – and highly controversial assumptions – about the identity of persons. For example, if you thought that the fact of a person's diachronic identity just consisted in facts about actions, you could then claim that, because the act of walking is necessarily extended, the person who is walking is therefore necessarily extended.

of time are always determined by the mental properties we attribute to them at each instant during that interval.

Soteriou presents the holistic conception as the solution to an apparent puzzle that can be brought out by our perception of motion. Soteriou's puzzle is generated by the following two assumptions:

Ass.1: A subject's awareness of a temporally extended occurrence will have the same temporal extension as the occurrence (2013: 97).[5]

Ass.2: 'if a subject is in a given state over a given period of time, then this is determined by … the fact that she is in that perceptual state at each of the instants that make up that period' (2013: 96).

According to Soteriou, Ass.1 'might be motivated by the temporal transparency of experience' (2013: 97). Experience is said to be 'transparent' in the sense that when we introspect, we cannot discover any features of our experience other than those the experience is *of*. Regarding the *temporal* features of experience, this amounts to the following: we are unable to identify any temporal location or duration of the experience other than what is presented *by* the experience. In other words, one's experience seems to run concurrently with its object. So when I am aware of an object's motion from location L1 at t1 to location L10 at t10, it seems that my awareness of this motion occupies that same interval – *viz.* t1 to t10. For example, when I observe the second hand of a clock sweeping from five past to ten past, it seems as though my experience occurs *as* the second hand makes its sweep; and it seems as though my experience lasts for the five seconds the sweep does.

When a subject is aware of an occurrence that unfolds over some interval, Soteriou supposes, they are in a 'perceptual state [that] *continues* to obtain *throughout* that interval

[5] I am presenting an abridged version of the puzzle. When a subject observes an unchanging state of affairs, Soteriou posits the claim that 'if you successfully perceive something enduring for an interval of time, t1 to tn, then you are aware *of* that thing *for* an interval of time, t1 to tn' (2013: 96). When the subject observes a *changing* occurrence, Soteriou posits the claim that '[i]f you successfully perceive an occurrence unfolding over some interval of time, t1 – tn, then there are sub-intervals of an interval of time, t1 – tn, *over* which you are aware *of* temporal parts of that unfolding occurrence' (2013: 97). It nevertheless follows from this latter claim that '[a]n occurrence that unfolds over time is something that you can only be aware of, as such, over time' (2013: 97). Either way, Soteriou thinks that a subject's awareness has the same temporal extension as whatever they are aware *of*.

of time' (2013: 94). So by Ass.1, my awareness of the object's motion from L1 to L10 spans some interval; and throughout this interval I am in a perceptual state of awareness of the object's motion from L1 to L10.

The problem is that this is in tension with Ass.2. According to Ass.2, the fact that I am in a state of awareness of the object's motion from L1 to L10 is determined by my being in that state of awareness at each of the instants during the interval of my awareness. But if this is the case, it is impossible for my state of awareness to span the interval of the object's movement: at, say, t2 I cannot be aware of the object's motion from L1 *to* L10, because the object hasn't moved to L10 *until* t10 (Soteriou 2013: 98). So my state of awareness of the object's motion from L1 to L10 cannot span the interval t1 – t10. This appears to contradict Ass.1.

Soteriou proposes that the way to resolve this puzzle is to adopt a relationalist view of perception; and once we do that we can reject Ass.2. For relationalists, 'there are phenomenally conscious states whose obtaining requires the obtaining of a relation of perceptual acquaintance, but which cannot be specified independently of that relation' (Soteriou 2013: 99). In the case we are currently considering, by Ass.1 I am aware of the object's movement from L1 at t1 to L10 at t10; this is my phenomenally conscious state. This phenomenally conscious state requires the obtaining of a relation of perceptual acquaintance, so the relationalist will say that the object's motion from L1 at t1 to L10 at t10 is an event I stand in that relation to. This relation will hold *over* the period during which the object moves – *viz.* from t1 to t10. And it is because this relation holds *over* this period that I can claim to be in a state of awareness of the object's motion from L1 to L10 at any moment *during* it. But we must then 'understand the claim as asserting that there are instants of time that fall within an interval of time over which [I am] aware of the movement of the object from [L1 to L10]' (2013: 106). In other words, although we can attribute these states of awareness, we can do so only in virtue of the fact that I am in a state of awareness *over* that period. As such, it is wrong to claim that my mental state over that interval of time is determined by my mental state at each moment during that interval. The atomistic conception has it the wrong way round.

I do not want to question whether the relationalist is able to solve the puzzle by rejecting Ass.2. What I do want to argue is that the puzzle can be solved by the atomist; and, further, that atomism enables a superior solution to the one suggested by Soteriou.

3.3 Responding to Soteriou's argument

The first objection I want to raise concerns the motivation of Ass.1. Ass.1 is, Soteriou says, motivated by the temporal transparency of experience. The transparency claim says that we 'see through' the experience, such that we only discover the temporal features of what the experience presents us with, rather than those of the experience itself. This, Soteriou says, is what motivates the claim that our awareness *of* a temporally extended event must itself be temporally extended – *viz.* Ass.1.

However, we need to be careful about making inferences from our temporal phenomenology to the temporal structure of experience. To paraphrase Moore (1903), I should not infer from the fact that I am having an experience *of* the colour blue that the experience itself *is* blue. Soteriou even acknowledges this point elsewhere, saying that one 'might think Moore's warning unnecessary, for there is little temptation to suppose that our mental episodes or states are actually blue in colour' (2013: 12). It is notable, then, that he *is* tempted to suppose that the temporal structure of an experience can be 'read off' from the phenomenology.

Indeed, evidence from psychology shows us that the following two claims cannot both be true together:

i) The temporal structure of a subject's mental life can be 'read off' from the phenomenology.
ii) A subject's awareness of a temporally extended occurrence has the same temporal extension as the occurrence [as Ass.1 asserts].

Take the 'oddball' effect for instance. If a subject is shown a series of identical stimuli, the first in the sequence will often seem to be longer in duration than the others (Rose & Summers 1995). Similarly, a stimulus will seem to be longer in duration than other stimuli of the same duration if it is differs from them in some other salient way (Pariyadath and Eagleman 2007).

Now, if the temporal structure of the experience matches that of the phenomenology – as (i) suggests – then, as the 'oddball' stimulus *seems* longer than the others, we will have to say that our awareness of the oddball stimulus does indeed span a greater interval than our awareness of each of the other stimuli. But this is inconsistent with (ii). The oddball

stimulus and the standard stimuli have the same duration. So if (ii) is true, my awareness of the oddball stimulus and the other stimuli will be equal in duration. As such, either (i), (ii), or both are false. So even if experience is temporally transparent – even if we can only ever discover the temporal features of whatever our experience presents us *with*, and can never discover the temporal features of the experience *itself* – this cannot justify the belief that one's awareness of an occurrences will have the same temporal extension as the occurrence itself. That is not to say Ass.1 is false. It's just to say that Ass.1 cannot be true for the reason Soteriou gives.

Let's assume for the sake of argument that Ass.1 is true, though; let's assume a subject's awareness of a temporally extended occurrence *will* have the same temporal extension as the occurrence. The second objection I want to raise concerns Soteriou's claim that the subject's awareness of the object's movement is a *state* of awareness. Without this claim, Ass.1 and Ass.2 won't generate the puzzle. Ass.2 is a claim about states, so the puzzle requires the subject's awareness described in Ass.1 to be a *state* of awareness in order to gain any traction.

Soteriou claims that when a subject is aware of an object's movement as it moves from L1 at t1 to L10 at t10, the subject is, from t1 to t10, in a state of awareness *of* that motion (*viz.* the motion from L1 at t1 to L10 at t10). The point I want to make is that it is impossible for an atomist to find this puzzling in the way Soteriou thinks we should. If the atomist accepts that the proffered mental state obtains they will have to deny Ass.2; or, if they accept Ass.2, they will have to deny that the mental state obtains. Either way, there is no puzzle. As such, we should only find motion perception puzzling for the reasons Soteriou provides if we have implicitly rejected the atomistic conception. Soteriou's argument cannot therefore give us reason to reject it.

Let's first suppose that the atomist is willing to say that the subject is in the proffered state from t1 to t10. They will hold that:

iii) From t1 to t10 the subject is in a state of awareness of the object's motion (*viz.* from L1 at t1 to L10 at t10).

But as they are an atomist, they will provide a reductive account of (iii). They will say that what makes (iii) true is that at t1 the subject is aware of the object at location L1, at t2 the subject is aware of the object at subsequent location L2, at t3 the subject is aware

of the object at L3, and so on. Different atomists will flesh this account out in various ways. For example, an atomist might think that the object of the subject's awareness is dynamic; or they might think that it is non-dynamic. Similarly, they might think that the subject is aware of a strictly instantaneous state of affairs; or they might think the subject is aware of some preceding interval.

Whatever the details, though, the atomist who accepts (iii) will, in providing this sort of reductive account, be set against Ass.2. Recall that Ass.2 is the claim that:

> if a subject is in a given state over a given period of time, then this is determined by […] the fact that she is in that perceptual state at each of the instants that make up that period. (Soteriou 2013: 96)

That is to say, if a subject is in a given state over a period of time, then they will be *in that same state* at each moment during that period. As we can see, this is what the atomist will deny in providing their reductive account of (iii). As such, if the atomist accepts the claim that the subject is in a *state* of awareness throughout the interval, they will not be able to accept Ass.2. The puzzle won't gain traction.

This position might encounter resistance. The objection would be that if a reductive account of (iii) is true then the awareness described in Ass.1 would no longer belong to the category of 'state'. The idea here would be that what determines the ontological category of a mental item 'is the way in which that item fills the relevant period of time – whether it persists through time, or occurs during the time, or obtains throughout the time etc.' (Steward 1997: 73). If a period of time is filled by a mental item that can be given the sort of reductive account we have just seen, then than mental item cannot be said to *obtain* over time but rather to *unfold* over time. However, mental items that unfold don't belong to the category of 'state'. States are 'non-dynamic' (Rothstein 2004: 6); they *obtain.* So the mental item described in (iii) would therefore be a state *by name only*.

The atomist might then instead reject (iii) and deny that the psychological property we attribute to the subject is a state. But if this is their position, Ass.2 becomes irrelevant. Ass.2 is a claim about states, and so it won't be saying anything about the property we are attributing the subject with. Again, the puzzle fails to get off the ground. So whether or not the atomist wants to call the mental item a state, they won't be troubled by Soteriou's puzzle. Hence, the puzzle will only *be* a puzzle to someone who has already presupposed the falsity of the atomistic conception; and, as such, Soteriou gives us no

reason to reject it. Further, the atomist is not forced to adopt a relationalist view of perception. Although relationalism is undergoing something of a renaissance at the moment, representationalism remains the most popular view.[6] As such, Soteriou's solution, in forcing us to reject representationalism and accept relationalism, is not easy to swallow. It would be much better to go with the solution that doesn't force our hand either way.

The atomistic conception therefore offers a better response to Soteriou's puzzle than holism. Indeed, by adopting the atomistic conception it is impossible to generate the puzzle in the first place, and further, we aren't forced to reject representationalism.

3.4 Phillips's argument

3.4.1 Fara's paradox

Now we can move on to Phillips's argument. Phillips makes his case in response to an ostensible paradox raised by Fara (2001). Fara's puzzle results from the two following observations:

Fara1: We experience constant motion

Fara2: Some positional changes are too small to be perceived

Fara1 is clear enough. An experience of constant motion is one during which the perceived object never appears stationary. Examples abound, such as when we watch a second hand sweep across a clock face. Similarly, there's no denying Fara2 – not without committing oneself to the incredible position that humans have infinite powers of visual discrimination. But by Fara2, there is a threshold positional difference below which we are unable to perceive change. Over small enough intervals, a constantly moving object will not achieve this threshold positional difference. Thus, over short enough intervals the second hand won't be seen to move. So an extended experience of motion will comprise

[6] According to representational theories of perceptual experience, perceptual experiences are intentional in that they are *about*, or represent, the world: 'perceptual states represent to the subject how her environment and body are. The content of perceptual experiences is how the world is represented to be' (Martin 1994: 464). There are many ways of unpacking this general idea, however, and little in the way of consensus as to the relation between the phenomenal qualities of experience – what it's like – and representation. See e.g. Macpherson (2014) for a recent discussion.

sub-intervals during which the perceived object appears stationary. But an experience of constant motion is one in which the perceived object never appears stationary. Fara2 precludes Fara1.

The solution Phillips proposes has two commitments:

Commitment 1: Experience has a temporal field within the bounds of which we perceive change.

Commitment 2: The content of the temporal field is purely determinable over short enough sub-intervals.

The terms Phillips employs here will be discussed shortly. By combining these two claims, we get the conclusion that '[e]xperiential contents are fundamentally temporally extended: they are the contents of temporal fields' (2011a: 823). Notice that this is a claim about the *content* of experience. It does not, by itself, establish the metaphysical priority of experiential *events*.

Some atomists think that, although experiential states are unextended, they have extended contents. They hold that experience is capable of representing extended periods of time, but that these representations do not themselves extend through time. These atomists are sometimes called retentionalists. It would appear, then, that the atomist *qua* retentionalist would be able to accept Phillips's solution to Fara'z puzzle. Phillips thinks so, at least:

> [T]he core components of the account below can be endorsed from a retentional perspective. From such a perspective the ideas below should be understood as focused on the contents of experience, as opposed to experience itself. The core of the account is that the dynamic content of our experience at short timescales is metaphysically dependent on the content of experience over longer timescales. (Phillips 2011a: 818)

Once we spell out what this would commit the retentionalist to, however, it becomes clear that Phillips's proposal is not an option for them. I explain why in the discussion of Commitment 1 below.

Against the retentionalist stands the extensionalist. Extensionalists think not only that experience represents extended intervals, but that these representations themselves extend through time. That is to say, they think that temporally extended phenomena are

represented in extended experiential events. Phillips is an extensionalist, and employs his argument within this extensionalist framework to argue for the holistic conception of experience. The idea is that if (as Phillips argues) experiential contents are fundamentally extended, and if (as the extensionalist claims) those contents unfold over time within extended experiential events, then (as the holist claims) the experiential events themselves will be fundamentally extended. Though one may wish to challenge this reasoning, I will assume it is correct. My focus will instead be on whether Phillips's two commitments provide a satisfactory solution to Fara's puzzle. By considering his proposed commitments in turn, we can see how his solution is intended to work.

3.4.2 Phillips's solution: Commitment 1

That experience has a temporal field is, Phillips claims, demonstrated by the fact that we are able to perceive the movement of some objects (or visual stimuli more generally) but not others. We can, for example, see the second hand sweep across the clock face, but the hour hand never appears to move. We only ever notice that it *has* moved. According to Phillips,

> What this shows is that there is an upper bound to the stretches of time over which we can directly apprehend complete events and processes. As a simple model, we can think of our experience as having a 'temporal field', and of there being limits to the extent of the field, say 300ms. (Phillips 2011a: 817)

The thought is that we perceive the world through a small temporal window of about 300ms. Since a positional change between indiscriminable positions appears as no change at all, the positions of a moving object must be discriminable within this 300ms window in order for the object to be seen to move. Only then is its positional change 'genuinely an object of experience' (2011a: 813). So either the positions are discriminable, in which case motion is represented as occurring over the 300ms duration, or else the positions are indiscriminable, in which case no motion is represented at all. This is the idea of the temporal field of experience.

Before moving on to Commitment 2, I want to show why the atomist *qua* retentionalist cannot accept this notion of the temporal field. It might not be obvious why this is the case. After all, the retentionalist claims that experience has a temporal field in the sense that experience is capable of representing temporally extended phenomena. This is the notion of what is sometimes called the 'specious present'.

As Phillips recognises, an atomist could be a retentionalist, meaning that they could hold that although the experiential state is momentary, its content is extended. And if its *content* is extended – if it is an experience *of* an extended event – then, assuming the experience is veridical, the *object* will be an extended event. So it seems that the atomist *could* accept Commitment 1, and hold that the object of experience is the 300ms change.

There are problems with this view, though. Let's say an observed object moves between two minimally discriminable positions *just* within the 300ms timeframe. If Phillips's notion of the temporal field is correct, the positional change will be a 'genuine' object of experience. But it would be what we may call a *borderline motion experience*. The positions are minimally discriminable, so if it moved any less in that 300ms, its positions would be indiscriminable and so it would not be seen to move. The problem is that, when understood within a retentionalist framework, this will either require positing an unrealistic delay-to-consciousness, or else it will require backwards causality.

For the retentionalist, experiential states are instantaneous, even though they can represent occurrences that take place over extended periods. So the retentionalist will posit an instantaneous experiential state (or a series of many such states) with the content of the object's positional change over a 300ms timeframe. But we may ask: when exactly does the (first) experiential state with this content occur? It must occur either prior to or upon completion of the 300ms positional change. If it occurs prior to the completion of the 300ms positional change, then the object of that experience will be an event that has yet to complete. In other words, we would be having an experience of an event that is partly in the future.[7]

To avoid this sort of backwards causality, the retentionalist is forced to hold that we experience the change upon its completion. This means that we would have been observing the moving object for (just under) 300ms before being able to see its motion. This is an unfeasibly long delay. Dennett (1991) considers a similar proposal in relation to the phi illusion. The phi illusion works by flashing two dots in different locations in quick succession. Subjects report seeing a single moving dot; the intervening motion is apparently 'filled in'. This result is problematic. Ruling out precognition, there are two natural lines of explanation. One is to say that the subject has a false memory of seeing the intervening motion: they believe they saw the motion, but they are mistaken. The

[7] Prosser (2016: 148) marshals similar considerations against the extensionalist view.

brain is merely telling the story that makes the most sense of what has just been experienced. This is what Dennett calls the 'Orwellian' hypothesis. The other is what he calls the 'Stalinesque' hypothesis. This time, instead of forming a false memory about what *was* experienced, there is a delay to consciousness. This allows time for both dots to be sub-personally registered and for the brain to then decide what to 'fill in'.

Over very short timescales, the Stalinesque hypothesis might seem fairly plausible. But, Dennett argues, all plausibility is lost when we consider Kohlers and von Grünau's (1976) version of the phi experiment, where there is 200ms between the spots:

> Suppose we ask subjects to press a button "as soon as you experience a red spot." […] Could it be that there is *always* a delay of at least 200msec in consciousness? No. There is abundant evidence that responses under conscious control, while slower than such responses as reflex blinks, occur with close to minimum latencies (delays) that are physically possible. After subtracting the demonstrable travel times for incoming and outgoing pulse trains, and the response preparation time, there is not enough time left over in "central processing" in which to hide a 200msec delay. (Dennett 1991:121-2)

If there's no chance of finding a spare 200ms for the Stalinesque delay, there's certainly no chance of finding a spare 300ms for the retentionalist trying to accommodate Commitment 1. Contrary to Phillips's earlier comment, then, Commitment 1 is unworkable within a retentionalist framework.

From here on, I will restrict my attention to considering the prospects of Phillips's argument within an extensionalist framework. This is what we are really interested in anyway; this is what, if it were to succeed, would undermine the conclusion from Chapter 2 (that questions of personal and experiential identity are empty).

3.4.3 Phillips's solution: Commitment 2

For the sake of argument, let's assume that Phillips is right about the temporal field. If a change is discriminable within the duration of the temporal field, we directly experience the object's movement. This tells us about the content of the temporal field as a whole. But we are still don't know how this can be possible given the fact that we have limited powers of perception. We don't know what to say about the content over sub-intervals of the temporal field. This is where the idea of *determinable* content comes in – Phillips's second commitment. To get an idea of what determinable content is, Phillips discusses an

analogous case (owed to Tye (2002)). If you look at a postage stamp with clear vision you will be able to make out the serrated edges; the content of your experience will be relatively determinate. But if your vision goes sufficiently out of focus, you will no longer see the serrations. Your experience will not represent the stamp's determinate shape. Nevertheless, the stamp will appear 'squarish'.

'Squarishness' is a *determinable* property: the edges of a squarish object could be straight, or fuzzy, or serrated, and so on, but 'squarish' remains silent as to these finer-grained details. Thus, we can say that the content of our experience of the out-of-focus stamp is in this sense determinable, but not determinate. It represents the stamp's determinable squarish shape, but does not represent the stamp's more determinate, serrated form.

Phillips claims that the same can be said of the content of the temporal field. We experience the clock hand's motion over the 300ms duration. But, due to our perceptual limitations, we are unable to determine any finer-grained details of the motion. There might be a change in velocity, for example, but such details are not represented in experience. Thus, the content of the temporal field is determinable, but not determinate:

> It is only over the whole period [of the temporal field] $\delta\tau$ that one has an experience of the hand sweeping out the angle $\delta\theta$. Moreover, it is not true that over any proper sub-period of $\delta\tau$ that one experiences, or is experiencing, the hand sweeping out some subinterval of $\delta\theta$: such tiny changes are beyond our powers of discrimination […] The most basic characterization of our experience during a sub-period of $\delta\tau$ is that we are experiencing the second hand sweeping out an angle $\delta\theta$. We do not experience this sweeping motion in any more determinate way [as] there is no more determinate truth about our experience at such timescales. (Phillips 2011a: 822-23)

It is true that what we are experiencing at an instant (or sub-interval) of the temporal field is motion. But what we are experiencing at an instant is not the hand's motion *at that instant* (or sub-interval). Instead, what we are experiencing at that instant is the hand's movement over a longer period; we simply experience the hand's motion '*continuing to unfold*' (Phillips 2011a: 821).

This shows us the way out of the paradox. It follows from Fara2 – that some changes are too small to be perceived – that it is impossible to perceive continuous motion. But it is important to distinguish between continuous and constant motion, at least as Phillips uses the terms. To perceive *continuous* motion it is necessary for the experiential content to be

determinate ‘all the way down’; we would have to perceive all the fine-grained details of the motion. To perceive *constant* motion it is sufficient that one experiences motion without any periods during which the object appears stationary. By Commitment 2, the fact that we cannot perceive continuous motion does not entail that we cannot perceive constant motion. Provided the content of the experience is purely determinable over short intervals, we will perceive the motion without all the fine-grained details. Hence, we can experience constant motion without perceiving continuous motion. Fara2 does not preclude Fara1.

It is impossible, Phillips thinks, to make sense of this if we take instantaneous experiential atoms as the most fundamental units of experience. Because the content of experience is purely determinable over short periods, what we experience at an instant is the hand ‘*continuing on* its motion through the larger interval’ (Phillips 2011a: 822). What this shows is that the content of experience is attributable at an instant only through reference to the content of experience over a longer period. Or, to put it another way, there is no truth about the content of experience at an instant that is not derived from what is true about the content over longer periods. In order to say *anything* about the content of experience at an instant, we have to first look to the extended experience during which that instant occurs. In short, we can only make sense of experience at an instant if we understand that instant as belonging to an extended experience. Thus, if content is purely determinable at short timescales, we have to accept that it is not instantaneous experiences but extended experiences that are fundamental. As this is the only way out of Fara’s paradox, we must reject the atomistic conception of experience.

3.5 Responding to Phillips’s solution

Before I start explaining where I think Phillips goes wrong, I want to quickly cover a possible objection I see being raised – but one that is, I think, unsuccessful. The objection is as follows. We are considering the question of whether extended experiential events are more fundamental than momentary experiential states. In his account, Phillips is discussing the *content* of experience. His account – if successful – would show that the *content* of a momentary experiential event depends on the *content* of the surrounding momentary experiential states. But, one might think, isn’t this just to make the same mistake we saw in the earlier Antarctica example? The mistake was to think that the dependency in question concerned the persons, rather than their positions in an ordering.

Of course the first and second *persons* don't form a fundamental whole; it is the first and second *positions* that do! Similarly, one might think, it would be a mistake to conclude that the momentary experiential *states* form a fundamental whole; the point is about their *contents*.

In Phillips's defence, though, it could be argued that there is an important difference in the case of experience. It is not essential to the existence of a *person* that they occupy a position in an ordering; whereas it is essential to the existence of an experiential event that it has content. Simply: an experience cannot be an experience unless it has content. To have an experience without content is to experience nothing at all. And so if the *content* of an experiential event depends for its existence on the contents of the surrounding experiential events, we can say that the experiential event *itself* depends for its existence on the surrounding experiential events. The same cannot be said of persons and their positioning in an order.

I am not sure how successful this response is. It depends on precisely what Phillips's argument establishes. If Phillips's argument establishes that a momentary (or very short) experiential state cannot have *any* content unless it is surrounded by other content-bearing experiential states, then the response is successful. But it is not at all clear that this should be the lesson to take from Phillips's argument. Phillips presents his solution as a way to explain how it is possible that we experience *motion.* Thus, what Phillips's argument actually establishes – if successful – is that a momentary experience of motion cannot be a momentary experience of motion unless it is surrounded my other momentary experiences of motion. And so if this demonstrates a metaphysical priority, the priority is that of extended experiences *of motion* over momentary ones. But not all the experiences we have are experiences of motion. It is not at all clear, therefore, that Phillips demonstrates the priority of extended experiences *per se* over momentary ones. That would require an extra step in the argument.

Perhaps we can supply that next step. Certainly, not all of our experiences are experiences of motion. But the lesson we can take from Phillips transposes to other sense modalities. For example, we are able to detect the change in the pitch of a constantly rising note; and yet we are unable to detect sufficiently small changes in pitch. Thus, Fara's paradox applies – and can be given the same sort of treatment by Phillips. We could do the same for taste, smell, temperature and so on. What I propose is that the content of experience is *never* totally static; and so long as it isn't, Phillips's conclusion will always apply.

I will now explain where I think Phillips goes wrong. His argument is, in short, that we can only escape the paradox by accepting his two commitments; and we cannot accept the two commitments unless we reject the atomistic conception of experience (that experience can be analysed in terms of fundamental momentary experiential states). My response will be as follows. Phillips's notion of the temporal field in Commitment 1 is too strong, but does get something right. We can therefore reformulate the first commitment, and when we do we see that the paradox can be avoided very easily and without invoking Commitment 2. As it does not follow from the revised first commitment that experience is fundamentally extended, Phillips's argument fails.

I do not disagree entirely with Phillips's first commitment. There is indeed a *sense* in which experience has a temporal field. The phenomenal difference between our experience of the hour hand and second hand does show us something important. But it does not establish quite as much as Phillips asserts it does. Phillips claimed that the difference between the two cases shows that 'there is an upper bound to the stretches of time over which we can directly apprehend complete events and processes' (2011a: 817). But this doesn't quite stack up. What the phenomenological difference shows is that, if an object (or visual stimulus more generally) we are observing moves a sufficient distance within a sufficiently short timeframe, we will experience motion. That much we can be sure of. But it is another thing entirely to claim that we experience the object's movement 'directly' within this timeframe. Once we entertain the thought that we experience the object's movement directly within the bounds of the temporal field – that the object's positional change from 0ms to 300ms is the object of experience – we slide uncritically to the idea that we experience motion over the *duration* of that timeframe – i.e. from 0ms to 300ms.

But this is too strong a claim. If a subject experiences motion over the duration of the temporal field, then even at, say, 1ms into that timeframe, it is true that they are experiencing motion. But how this could be true is unclear, to say the least. We are all agreed that our powers of perception are limited. Thus, it should be uncontroversial that, within, say, 1ms of observing a moving object, it is physiologically and neurologically impossible to perceive the object's motion.[8] Sure, Phillips can appeal to his second

[8] We have a limited number of light-receptor cells in our eyes, and so a stimulus has a minimum distance to move in order to stimulate a different group of cells. And then once a new group of cells has been stimulated, we have to wait for the signal to work its way to the brain and then be processed. This all takes time.

commitment (*viz.* that the content is purely determinable at such timescales) in an attempt to explain how this might work. But this appeal just obscures the problem, not answer it.

Firstly, such an explanation looks hopelessly circular. Phillips claims it is true that a subject is experiencing motion at 1ms. But because the content at such timescales is purely determinable, this is true only in virtue of the fact that the subject is experiencing motion over the *whole* of the temporal field. But wait a second (pun not intended). How can it be claimed that the subject experiences motion for the duration of the temporal field? That's precisely the point we are challenging. Even if we allow, for the sake of argument, that experiences *are* fundamentally extended, we are still waiting for an explanation of how the extended experience can have the extension Phillips is claiming it to have: how can the experience of motion begin so soon after first looking at the moving object? No explanation for this is given, other than what is asserted in the first commitment: that the movement is directly perceptible within the bounds of the temporal field. So in order to lend plausibility to the first commitment, we look to the second commitment; but the second can't save the first unless the first is assumed to be true.

Ultimately, though, this is beside the point. Let's suppose the two commitments *can* supply a non-circular explanation. Unless he wants to assert that it *is* neurologically possible to detect an object's motion within such a short timescale, Phillips's solution seems to have the consequence that we can truthfully say of a subject's experience something that, neurologically speaking, cannot be true of it.

In Phillips's defence, one might argue that when we first look at a moving object there does not *seem* to be an initial period of stasis. When we are about to cross the road and look up at a moving car, the car does not appear stationary for a brief time, before suddenly jolting into motion. This is quite true. But it does not follow that during the initial period we must be experiencing it as moving. The car will only appear stationary if it is represented in experience as stationary. It does not appear as stationary, so clearly it is not represented as stationary. But it does not follow that during that period it must therefore be represented as *moving*. The details of the car's motion could simply be omitted from the representation. And, if it is neurologically impossible to detect the motion during that initial period, we must assume this is the case. So, during the initial period, although we do not experience the car as moving, *we don't experience it as not moving either.* To use Phillips's own idea, we can say that our experience of the car will be determinable but not determinate. Unlike the car's motion, it is, I assume,

neurologically possible to perceive its general size and colour almost immediately. But because we cannot yet perceive its motion, the representation of the car will have many potential determinates. The car might be stationary, or it might be moving; and if it is moving, there are many potential speeds it could be moving at. Our experience of the car will be detailed *to a* degree, but indeterminate with regards to such details. By that I mean to say that it will be indeterminate as to what is represented; not that the representation will be a representation of indeterminacy.

An opponent might argue that this explanation is inadequate. I have explained why the car does not appear stationary. But not only does the car *not* appear stationary when we first look at it, it *does* appear to be moving. And if it *appears* to be moving, then it must be *represented* as moving.

I think this response is wrong; I think it *cannot* be the case that the car appears to be moving. The only way this could be true is if we posit a delay-to-consciousness, which would allow time for the brain to work out that the car is in fact moving. But, as Dennett's response to the Stalinesque hypothesis we looked at earlier shows, it doesn't look like this move will stick. The only option left is to deny that movement is represented when we first look at the car. If you think it *seems* like movement is represented the *moment* we look at the car, then you must accept that this is a conclusion the brain has come to *once the motion is detected.* After all, if you claim to have perceived something that it was *neurologically impossible* to have perceived at the time you are claiming to have perceived it, it is, I submit, safe to say you're mistaken.

Perhaps Phillips would be willing to accept that motion is not experienced *right away*: when the subject first directs their attention to a moving object, there is an initial period of perceptual stasis. Only once it is neurologically possible does the extended experience of motion begin. This would be a step in the right direction, I think. But it does make a nonsense of the first commitment. The first commitment states that we perceive an object's movement within the bounds of the temporal field. But if there is a sub-interval of the temporal field during which we do not perceive motion, this can't be the case. It can't be the case that an object doesn't appear to move when we are directly perceiving its movement. Commitment 1 should, I conclude, be rejected. Experience does not have a temporal field through which we directly perceive movement.

Nothing in what has been said demonstrates the second commitment to be problematic, however. From what has been said, it could well be the case that the experience of motion – once it actually gets going – *is* purely determinable at short timescales. With a slight adjustment, Phillips's model might be salvageable.

Phillips thinks that experiential content is determinable at short timescales below the duration of the temporal field: we can only state what is represented during any sub-interval of that duration in terms of what is represented over the whole duration. Well, now we are saying that the experience of motion doesn't unfold over the *whole* duration of the temporal field. That would be neurologically impossible. But, presumably, it *will* unfold over a sub-interval (B) of the period we are referring to as the 'temporal field', following an initial period of perceptual stasis (A):

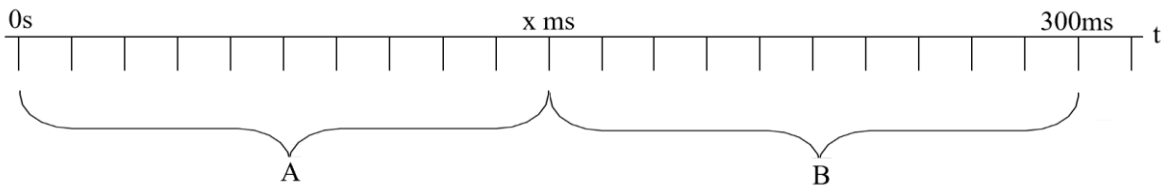


Figure 4: the temporal field

As such, Phillips could just say that the most determinate fact about what is represented in experience is a fact about what is represented over the duration of B. In other words, he could just argue that the starting point for the experience is later than we have been so far supposing. He could still maintain that at sub-intervals of B, content is purely determinable. And if the content is purely determinable, we have a strong case for regarding extended experiential events as fundamental (just slightly less extended than Phillips first claimed).

The question now is: do we have reason to believe that the content of the sub-intervals of B is purely determinable? I don't think so. My thinking is this: Phillips's argument was that we have to accept his two commitments in order to escape Fara's paradox. But, as I will show in the next section, now that the first commitment is out the window, a much simpler solution is available. It *might* be the case that content is purely determinable over very short intervals, but we can escape the paradox without supposing so. As such, Phillips's second commitment is unmotivated, obsolete and superfluous.

3.6 A better solution to Fara's paradox: the Simple Model

As argued, we should reject Phillips's first commitment. We do not directly experience movement over the duration of the temporal field. We should allow for an initial phase of perceptual inactivity. The motion phenomenology then 'kicks in' when a sufficiently great positional difference is observed in a sufficiently short timeframe. This borrows some of Phillips's idea of a temporal field but leaves the rest. It takes the idea that we experience of motion as the result of a observing a change that occurs within a certain timeframe. But it leaves out his idea that the object's movement is perceived 'directly' within that period. As we saw, this latter component gets us into trouble. The idea we are left with is:

Revised first commitment: A subject enjoys (non-illusory) motion phenomenology when and only when an observed object is in a sufficiently different position from a sufficiently recent position it was observed in.

The revised first commitment does not say that the positions must be *consciously* discriminable, nor does it state what the timeframe is. These details will be fleshed out soon.

Once we equip ourselves with the revised first commitment, a straightforward solution to the paradox presents itself. To illustrate, let's say that we begin observing a constantly moving object at t0, which achieves the minimum sufficient difference for motion phenomenology at t10. Also assume that this is within the timeframe necessary for motion perception. At t10, we would experience motion:

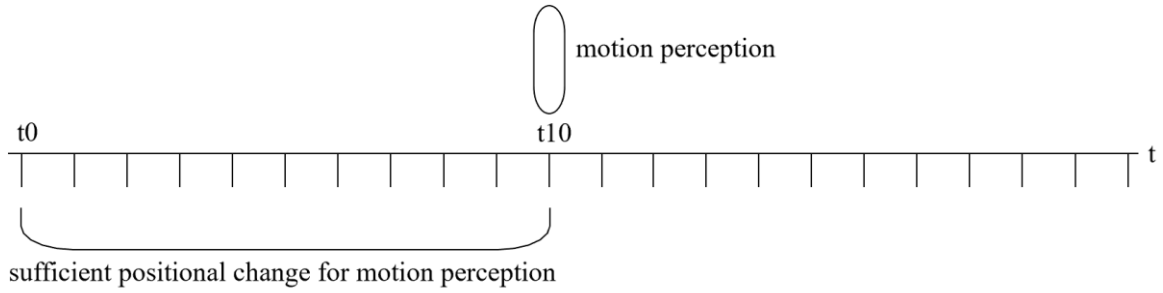


Figure 5: the simple model i

At t11, the position of the object will be indiscriminable from its position at t10. This, according to the paradox, means that we will not perceive motion. This is wrong, though. It is wrong because although at t11 the position observed is insufficiently different from the position observed at t10, it *is* sufficiently different from its observed position at t1:

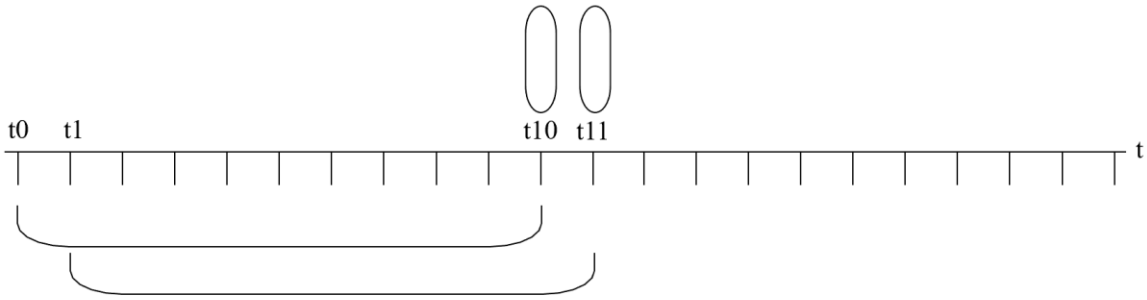


Figure 6: the simple model ii

And so, we can see how from t10 onwards we will experience motion phenomenology. And it is clear that no appeal to determinability or extended experiences is needed. All we need is the revised first commitment. Call this the *simple model*.

It might be objected here that the timings I am suggesting are unrealistic. Phillips claimed that we experience motion when a positional change is discriminable within 300ms. So doesn't my model say that we will be looking at the object for 300ms before we start to experience motion? This, one might introspect, is too long. This is not a problem for the simple model. 300ms is not my figure. Consider again what I have called a 'borderline motion experience'. This is an experience of motion that we are only just able to perceive. We can imagine an hour hand of a clock so large that we are able to detect movement at its tip. If the hand was any shorter, and the movement would be imperceptible. If it *is* the case that the motion phenomenology kicks in before 300ms, then *that* is the timeframe we are looking for: that is the timeframe within which the positions of a moving object must be sufficiently different to generate motion phenomenology.

At this point Phillips would object. We experience motion some time before 300ms, yes, but the positions occupied by the object would only be discriminable *at* 300ms. So anything less than 300ms and the positions would be *indiscriminable*. But if the positions are indiscriminable over that sub-interval, we cannot perceive the positional change over that sub-interval; and if we can't perceive the positional change over that sub-interval, then we cannot have an experience of motion over that sub-interval. The experience of

motion, therefore, must happen over the longer period where the positions *are* discriminable.

I do not accept this. I don't think the positions need to be *consciously* discriminable in order to perceive motion. Admittedly, it will usually be the case that, when we perceive the movement of an object, its positions will be consciously discriminable. But it is a mistake to consider conscious discrimination a necessary condition for motion phenomenology.

As I see it, Phillips is running together two importantly distinct neural mechanisms here. One mechanism computes (or gives rise to) motion *per se*, the other computes (or gives rise to) positional discrimination.[9] There are several examples of these mechanisms coming apart. In the 'waterfall illusion', for example, a scene looks to be moving despite there being no apparent changes in the position of any of its visual features. More striking is akinetopsia, a condition that renders the subject unable to detect motion, despite an awareness of positional change. Subjects report seeing the world as a succession of stills (see e.g. Zeki 1991). Similarly, the 'fine grain motion illusion' suggests that we can have an experience (as) of motion between two points too close to consciously discriminate (Thorson, Lange, and Biederman-Thorson 1969).[10] More on that soon.

If we can truly say of our experience that we are experiencing motion *before* the positional discrimination (as Phillips says we can), then surely the natural conclusion to make is that the mechanisms for motion experience is more sensitive than the mechanism for positional discrimination: we are able to consciously detect the motion of an object over smaller distances than we are able to consciously discriminate positional change.

---

[9] The strange thing is, Phillips acknowledges as much when he says that 'empirical work on motion perception suggests that our visual systems can respond to motion *per se*, as opposed to simply detecting the change of position of features over time […] we may see motion without any awareness of position […] I abstract from these empirical issues here, and provide the broad structural outlines of an account of constant motion perception' (Phillips 2011a: 820-1). I don't see how he can abstract from these issues as he does.

[10] A caveat: although motion is being computed in this experiment – that it, the subjects have an experience *as of* motion – it is nevertheless an illusion. They are being presented with two stationary dots separated by a brief interval, rather than a single, continuously moving dot. What I really want to establish is that motion is *veridically* computed over smaller distances than can be discriminated. The required experiment would very similar, but instead of flashing two dots 50ms apart, a single dot would be displayed that takes 50ms to travel between the two positions. I am assuming that the result would be the same – that motion would be perceived in the absence of discriminable positional change. I haven't managed to find such an experiment.

The fine-grain motion illusion lends some support to this idea. In this experiment, two dots are flashed on a screen in quick succession. These dots are so close together that when the time between flashes is short enough, the two dots appear as a single *unmoving* dot. Subjects are unable to discriminate between the two positions. Similarly, when the time between flashes is long enough to notice, the second dot appears to be in the same place as the first. The positions, it is natural to conclude, are, at the conscious level, indiscriminable. But when the time between flashes is around 50ms – between the two extremes – the subjects report seeing a single *moving* dot. This experiment arguably shows that motion can be perceived between points too close to consciously discriminate between.

Whether the fine-grain motion illusion establishes motion perception to be more sensitive than positional discrimination could be contended. It could be argued that when the time between flashes is 50ms the positions *become* consciously discriminable, whereas when the timing is different they are indiscriminable. This move would ensure that we cannot perceive motion without positional discrimination: whenever we perceive motion we are able to discriminate positions. I do not want to get too stuck into this potential reply. But it should be noted that it contradicts the established orthodoxy. And besides, given that we are talking about two distinct neural mechanisms here, it seems odd to insist that they would both be *equally* as sensitive. It is far more plausible to allow that, of any two neural mechanisms that process sensory information, one will be more sensitive (quicker to respond etc.) than the other. And if it is allowed that one will be the more sensitive, and it is also intuited that we are able to perceive motion before we are able to discriminate positions, then it is going to take some arguing to convince us that it is not the motion mechanism that is the more sensitive of the two.

Ultimately, though, the simple model is flexible enough to work with whatever the neurological facts are. If the delay to motion perception is less than the delay to positional discrimination, no problem. All we need to explain this fact is the revised first commitment. If they are the same, no problem. Again, all we need to explain this fact is the revised first commitment. But if they are the same, a proponent of the simple model would reject reports that motion is experienced before positional discrimination. If a subject's reports conflict with the neurological facts, the subject must be mistaken. But either way, whatever the neurological facts are, there is no need to appeal to the two commitments Phillips does. All we need to escape Fara's paradox is the revised first

commitment, and *no mention of determinability*. The paradox therefore gives us no reason to reject the notion that experience can be analysed in terms of momentary (or very short-lived) 'atoms'. I conclude that Phillips does not succeed in establishing that experiences are fundamentally extended.

3.7 The Cartesian Theatre objection to the Simple Model

I will now consider what I anticipate will be the main objection to the simple model. The objection would be that the simple model assumes what Dennett calls 'Cartesian materialism', which regards experience as akin to a sort of internal motion picture, with us – the subjects – as the audience:

> Cartesian materialism is the view that there is a crucial finish line or boundary somewhere in the brain, marking a place where […] *what happens there* is what you are conscious of. (Dennett 1991: 107)

Our sense-organs collect data which is encoded in the brain before making its way to the 'finish line' whereupon it is transduced into consciousness. The 'finish line' need not be taken too literally. Although a naïve Cartesian materialist might think there is a specific *location* for the neural-phenomenal transduction, the finish line could instead be thought of by the Cartesian materialist as the culmination of neural processing of a special sort. The relevant point to our discussion is that on the Cartesian materialist view, there is a determinate matter of fact about what information has been transduced and is therefore experienced by the subject at any particular moment. By assuming Cartesian materialism we imagine that we could, if you like, *stop* the film reel, and inspect the frame being shown *at that moment*. We can always ask 'at what moment did that particular content become conscious?'. But this question, the argument goes, betrays a misunderstanding:

> […] being an item in consciousness is not at all like being on television; it is, rather, a species of mental fame. Almost literally. Consciousness is cerebral celebrity – nothing more and nothing less. Those contents are conscious that persevere, that monopolize resources long enough to achieve certain typical and "symptomatic" effects – on memory, on the control of behavior and so forth. (Dennett 1993: 929)

There are two related criticisms that can be brought out here. Firstly, just as it makes no sense to ask at which moment a celebrity acquires fame, it makes no sense to ask at which

moment a particular content becomes conscious. And yet, in saying that that motion is experienced after precisely x ms (at t10), the simple model provides an answer to this question. As such, the simple model should be rejected for assuming Cartesian materialism; there can be *no* precise time at which the experience of motion begins.

The simple model could also be accused of implicitly assuming Cartesian materialism in that it assumes there is a definitive fact of the matter about what is experienced at an instant. When we reflect on our experience, we imagine there is some canonical version of the phenomenal events to which we have just been subject. And by constructing this narrative it seems plausible that there is a fact of the matter about what was being experienced at any particular instant. But, the objection goes, this narrative – and therefore the narrative-at-an-instant – is a fiction; there is *no* canonical version of the phenomenal events. There are at any one time many different versions, all competing for influence (fame) throughout the brain.

I will now respond to these objections. I do not think that the first objection is too much trouble for us. Recall that Figure 6 shows the subject's experience of motion determinately beginning at t10. The objection here would be that the contents of consciousness aren't the sorts of things that we can say have determinate start points. Neural influence isn't an on/off affair: it *spreads*, and is a matter of degrees. Thus, there can be no moment when the experience of motion becomes determinately conscious. This might be true. But if it is true, we can build it into our model. Instead of the model positing an on/off experience of motion, it can chart the increase of neural influence with time. Where our current model shows the determinate beginning of the motion experience, we can take this as the point at which the neural influence is sufficient to prompt a verbal report. There would, I assume, be a short duration between retinal excitation and subsequent impulse-processing by the relevant motion-detection module in the brain. And so, if motion detection is what we are modelling, our model would still show an initial period of inactivity. But now there will be an intermediary period of neural propagation:

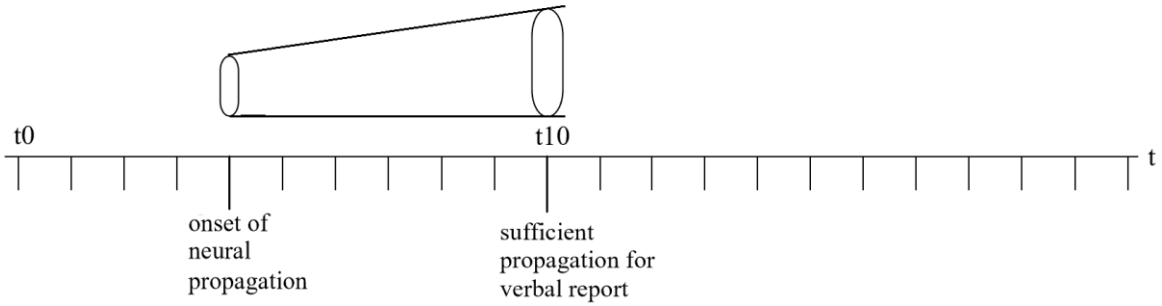


Figure 7: the simple model iii

With respect to the first criticism, then, the Simple Model does not need to assume a Cartesian materialist conception of consciousness. And, I think, if we take the simple model as modelling neural propagation (which is, according to Dennett and others, *all there is* to consciousness) rather than the broadcastings of the Cartesian Theatre, we have a way to respond to the second objection. The second objection was that the idea of a canonical narrative-at-an-instant is a fiction, because there are many different competing versions at any one time. Well, if we are now modelling the propagation of information from specific modules, there *will* be a 'canonical version' at an instant. Not of the stream of consciousness *per se* – because there is no such thing as a 'canonical version' (or so we are assuming for the sake of this objection). But there *will* be a single, determinate matter of fact about what sort of effects the activities of a particular module of the brain will be having on other parts of the brain; about how far the information has spread and its efficacy in terms of verbal reports, behaviour and so forth. And, moreover, there will be a fact of the matter about these effects *at an instant.* We don't need to assume that there is a Cartesian Theatre to accept this. If we truly embrace the idea that there is nothing more to consciousness than neural propagation, then we *can* model consciousness – or at least *aspects* of consciousness – with the simple model.

Of course, it may well be the case that there *is* more to consciousness than neural propagation. There may be a processual 'finishing line'. Whether or not there is, the Simple Model is flexible enough for us not to worry. The simple model shows us a way to escape Fara's paradox without appealing to the notion of determinable content. Phillips's argument for the holistic conception of experience is therefore unsuccessful. The atomistic conception of experience is vindicated.

3.8 Summary

Neither Soteriou nor Phillips have presented compelling arguments in favour of the holistic conception of experience. On the contrary, their own holistic solutions to the so-called puzzle cases they raise run into deep problems, and the puzzles themselves can be dealt with far more effectively by the atomist. We saw that Soteriou's solution, in relying on the temporal transparency of experience, falls foul of empirical evidence; and, further, that the puzzle he attempts to solve can only *be* a puzzle for the holist. Phillips's solution to Fara's puzzle fared no better. It was shown that his notion of a temporal field central to his account, although ostensibly plausible, not only had unacceptable neurological consequences but also ran together two importantly distinct neural mechanisms. Once the appropriate corrections are made, a straightforward atomist solution to the paradox becomes available. The Neo-Parfittian claim that it is an empty question whether experiences have temporal extension is safe. We can put this claim to use against the A-theorist's argument from experience.

Chapter 4: The Counter-Argument from Experience

I have argued in the previous chapters that there is no metaphysically non-trivial way of unifying the different momentary experiential states that we uncritically suppose comprise a subject's stream of consciousness. While we can *describe* two experiential states as belonging to a single experiential event, it would be just as metaphysically accurate to describe them as not. What I now want to do is deploy that conclusion against the A-theorist's argument from experience. Recall that this is that argument that:

(A1) We have experiences as of A-theoretic change.

(A2) The best explanation for (A1) is that A-theoretic change occurs in reality.

(A3) Therefore, A-theoretic change occurs in reality.

In this chapter I respond to (A2), assuming for the sake of argument that (A1) is true.

The thought behind (A2), I take it, is that it makes more sense to suppose that our experiences as of A-change are veridical than it does to suppose that they are non-veridical. After all, the traditional B-theoretic response to temporal experience is to dismiss it as illusory; a move the A-theorist wants to block. If we can show that our experiences as of A-change are non-veridical, then, we will have sufficient reason to reject (A2) from the argument from experience.

I will now argue that even if A-change was an objective feature of reality, it couldn't be an object of perception. The argument in broad strokes is that we are not justified in believing that A-change is an object of perception given that it is an empty question whether experiences have temporal extension. The argument – call it the *counter-argument from experience* – looks like this:

	(B1)	The A-theory can be evidenced by experience only if A-change is an object of experience.
	(B2)	A-change cannot be an object of experience if our perceptual awareness of extended events is derivative.
	(B3)	Our perceptual awareness of events is derivative if it must be accounted for in terms of momentary experiential states.
	(B4)	Our perceptual awareness of events must be accounted for in terms of momentary experiential states.
(B3), (B4)	(B5)	Our perceptual awareness of extended events is derivative
(B2), (B5)	(B6)	A-change cannot be an object of experience.
(B1), (B6)	(B7)	The A-theory cannot be evidenced by experience.

The argument is valid, but some of the premises will take some explaining. This is due in part to the terminology, which will be made explicit as I go.

4.1 Justifying (B1)

(B1) is reasonably straightforward. Changes – the sort we are interested in – are temporally extended events involving, minimally, two distinct (or incompatible) states of affairs, P and Q.[1] Because the B-theorist denies the existence of a moving present, they understand the change from P to Q merely as tenseless 'at-at' change: P obtains at t1 and Q obtains at t2 (Russell 1903). The A-theorist demands more from the notion of change. Because for the A-theorist what is present is constantly changing, not only does P obtain at t1 and Q obtain at t2; P is *replaced* by Q as the presently existing state. We can thus say that A-theoretic change, or A-change, is understood as *Q replacing P*.

If experience provides evidence of the A-theory as the A-theorist's argument from experience claims, it must be the case that we have veridical experiences of *Q replacing P*. If we don't have veridical experiences of *Q replacing P*, either we will have non-

[1] The sort of change we are interested in is diachronic change, so I ignore the sort of change (if one can even call it that) involved in spatial variation – e.g. the change between the hot and cold ends of a poker.

veridical (illusory or hallucinatory) experiences as of *Q replacing P*, or else we will have no experiences as of *Q replacing P* whatsoever.[2] Either way, experience could not be taken as evidence for the A-theory.

Now, in order to have a veridical experience of A-change, it must be the case that A-change is an object of experience. One cannot have a veridical experience of something that is not the object of experience. To see why this is the case I must first explain what I mean by 'object of experience'. This term might initially lead to some confusion. One finds terms like 'direct object of experience' and 'direct object of perception' populating the literature. As a loose definition:

> [P]erceiver P *directly* perceives an object O if P perceives O without perceiving any intermediary I. P would be perceiving an intermediary I if, as things are, it is only in virtue of perceiving I that P perceives O. (Dancy 1985: 144, emphasis added)

Different theories will take different stances on what these direct objects are. For instance, a sense-datum theorist will hold that the direct objects of perception are sense-data; whereas a naïve realist (or relational theorist, or object theorist) will hold that they are the mind-independent objects 'out there in the world' that experience presents us with. When I speak of 'objects of experience', I too am referring to mind-independent objects. But I am not using 'object' in the sense outlined by Dancy. I am not committing myself to any view about what the *direct* objects of perception are. There may be perceptual intermediaries (sense data, etc.), or there may not be. The most I mean to commit myself to is a loosely realist view of mind-independent reality: mind-independent objects exist, and play a causal role in our perceptions of them.

That is not to say that I am assuming the causal theory of perception to be true. Although the causal theory of perception is controversial, it is generally agreed even by those that oppose it that Grice (1961) did succeed in at least stating a *necessary* condition for perceptual experience: for subject S to perceive object O, O must causally influence S. There is a causal chain starting at a mind-independent object that leads to the subject's sense organs and which ultimately results in some particular kind of mental activity. This much is uncontroversial, and I assume so much to be true in my argument. What is

[2] When I talk of 'veridical experiences of *Q replacing P*' I mean veridical *with respect to Q replacing P*.

controversial is whether or not this is the object with which we are in direct conscious contact with. A direct realist/ relationalist will think that it is, whereas a sense datum theorist will think we are in direct conscious contact with some mental item. Similarly, a sense datum theorist will identify the perceptual experience with the mental activity at the 'end' of the causal chain, whereas a relationalist will hold that the perceptual experience is the *relation* between that mental activity and the mind independent object. These disagreements are irrelevant to my aims. When I use the term 'object of experience', I am merely referring to the mind-independent object that the perception is *of*; the thing in the world that, under some appropriate description, directly explains the qualitative character of the experience (in veridical conditions). Thus, it will suffice for our purposes to say that an object of experience is (a) mind-independent, (b) what the experience is of, and (c) causally responsible for the experience of itself.[3] The existence of these sorts of objects is agreed upon by anyone who holds a realist view of mind-independent reality. Whether or not this is the 'direct' object of experience (or a constituent of the perceptual experience) is not relevant to the current argument. Hence, if we veridically experience A-change, A-change must be the object of experience.[4] It then follows that the A-theory can be evidenced by experience only if A-change is an object of experience.

4.2 Justifying (B2)

(B2) claims that A-change cannot be an object of experience if our perceptual awareness of extended events is derivative. This can be unpacked as the following two claims. Firstly, A-change cannot be an object of experience if we cannot be perceptually aware of extended events. And secondly, A-change cannot be an object of experience if that awareness is derivative. I will discuss these in turn.

The object of experience was defined above as the object that lies at the beginning of the causal chain that results in, or is constitutive of, the perception of that object. So the object of awareness is by definition that which we are perceptually aware of. If we aren't

[3] The difference between a hallucination and an illusion then turns on whether the representation of an object has an object of experience, in the defined sense. An illusion has, whereas a hallucination lacks, any corresponding causally responsible mind-independent object. The difference between a veridical perception and an illusion then turns on the accuracy of the representation.

[4] I assume throughout that to perceive X is to veridically experience X.

perceptually aware of something, that something can't be the object of experience. Hence, if we can't be perceptually aware of temporally extended events, then the objects of experience can't be temporally extended; the only objects of experience we would be capable of detecting would be durationless. A-change (if it occurs) is a temporally extended event. It then follows that A-change can be an object of experience only if we can be perceptually aware of temporally extended events. This secures the first claim.

It might be objected at this point that the objects of experience can be extended even if our perceptual awareness is only ever of states. After all, everyday objects persist – that is to say, they have temporal extension – and it might be held that we are capable of detecting these objects insofar as we are capable of detecting instantaneous states of these objects. So it seems that the objects of our experience *can* have temporal extension, even if we can only ever be perceptually aware of durationless instants.

It may be true that our awareness of persisting objects does indeed come from our perceptual awareness of the states they are comprised of. But if our perceptual awareness of objects is *restricted* to their states, then what awareness we can have of these persisting objects is built up – derived – from such perceptions. We would be aware of persisting objects *in virtue of* perceiving their states.

This leads us to the second claim – that A-change cannot be an object of experience if our perceptual awareness of events is derivative. 'Derivative' and 'non-derivative' first need to be defined. As I am using the terms, we are *derivatively* aware of an event if our awareness of that event is derived from an awareness of states of (or during) that event.[5] We are *non-derivatively* aware of an event if our awareness is not derived from states of that event; if our awareness is of the event *per se*.

Let's assume for the sake of argument that change is A-theoretic. A-change events have the metaphysical structure of *Q replacing P*. P and Q, we may suppose, obtain within adjacent intervals. We may suppose this to be true whether the change is continuous or discrete. If it is discrete, the change will be a 'jump' from one state of affairs to another between adjacent intervals. With continuous change, one might object, the same cannot be said. The thought goes that what makes a continuous change continuous is the fact that

[5] Again, my use of 'event' and 'state' is idiosyncratic. When I say 'event', I merely mean something with temporal extension. When I say 'state' I am talking about something without temporal extension.

it does not jump from one state to another between adjacent intervals. The change from P to Q will be continuous precisely because there is an intermediate interval between fully-P and fully-Q during which the state of affairs phases from one to the other. This is not the full story, though. If there is a continuous change from P to Q there will be an interval during which fully-P is true, followed by an adjacent interval during which fully-P is no longer true. Not-fully-P is a distinct state of affairs from fully-P; we could call fully-P 'R' and not-fully-P 'S'. And so we would have a change from R to S between temporally adjacent intervals.[6] Thus, change is always between distinct states of affairs in adjacent intervals whether the change is discrete or continuous. We can therefore think of A-change as necessarily involving the replacement of one state with another.[7]

If our awareness of events was merely derivative we would be aware of the event 'Q replacing P' in virtue of being aware of its states. We would be aware of P, and of Q, and of the ordering of those states. But this does not amount to an awareness of A-change. It amounts to an awareness of change as described by the B-theorist. For the B-theorist, change is simply an ordering of states of affairs. There is no *replacement* of states. And it is precisely this process of replacement that would be beyond our perceptual grasp, were our perceptual awareness of events merely derivative. Anything that is beyond our perceptual grasp cannot be an object of experience. Hence, A-change cannot be an object of experience if our perceptual awareness of events is merely derivative.[8]

[6] The same reasoning can be applied for any intervening state between R and S.

[7] One might worry here this account of change has the consequence that *proper* continuity is impossible. I think this worry would be misplaced. Yes, all change involves change from one state of affairs to a distinct state of affairs. And so it might be said that in this sense even continuous change is discrete. However, there a crucial difference between discrete and continuous change. With a continuous change, between any two states of affairs there will be a further, intermediary state of affairs. This is what 'true' continuous change is. Change is discrete when it is between two states of affairs with no intermediary state.

[8] I do not know how to show that A-change *could* be an object of experience if our awareness of events was *non*-derivative. But it seems plausible to allow that it would be possible. *Q replacing P* is an extended event, and so it seems reasonable to assume that, if we could be non-derivatively aware of that event, we would be aware of the replacement. This does not matter for the argument. All I need to show is that A-change cannot be an object of experience if our awareness of events is merely derivative.

4.3 Justifying (B3)

(B3) claims that any perceptual awareness of extended events must be derivative if experiential states are momentary. If the A-theorist rejects (B3), they will hold that we can be non-derivatively aware of extended events even if experiential states are momentary. They will hold that our awareness is not derived from the perception of states during that event; they will hold that our perceptual awareness is of the event *per se.*

We are assuming the broadly realist view that mind-independent objects of perception exist and that they play a causal role in our perceptions of them; when we are perceptually aware of an object, we are causally connected to it. The object of perception in question is an event. To be non-derivatively aware of an event – that is, to be perceptually aware of the event *per se* – our experience must be causally connected to that event *per se* and not merely to the states during that event. The causal connection between our experience and the event must be non-derivative. If it wasn't, the objects of experience would be states, and our awareness of the event would be derived from those states. Our awareness of the event would be merely derivative. Someone who rejects (B3) therefore holds that experiential states can be non-derivatively causally connected to events.

What I want to first establish is that, if the A-theorist is to hold that experiential states can be non-derivatively causally connected to events, they will be committed against the principle of *temporal locality*. Once this is done I will go on to show why this commitment spells disaster for the A-theorist.

Temporal locality can be expressed in slogan form as 'no unmediated action at a temporal distance'. The idea is that a cause and its effect cannot be separated by a temporal distance without there being some causal intermediary, or chain of intermediaries, linking the two. Someone who rejects (B3) is committed against temporal locality; they must accept that causal relations can hold over temporal distances in the absence of any intermediaries. In other words, they must accept temporal *non*locality. The reason for this is as follows.

Let's say that event *A* (A-change) occurs from t1 to t3, at which time (t3) experiential state *B* obtains. As discussed above, rejecting (B3) means the A-theorist must hold that the causal connection *B* has with *A* is non-derivative. So *B* must be causally connected to *A per se*; that is, *B* must be connected to *all* of *A.* That being so, there will be a causal connection terminating at t3 (where *B* occurs) and stretching back to t1 (where *A* begins).

But of course, *B* doesn't exist at t1. So it seems that we have an instance of action at a temporal distance – *viz.* between t1 and t3 (and also between any intermediate time and t3 – see Figure 8).

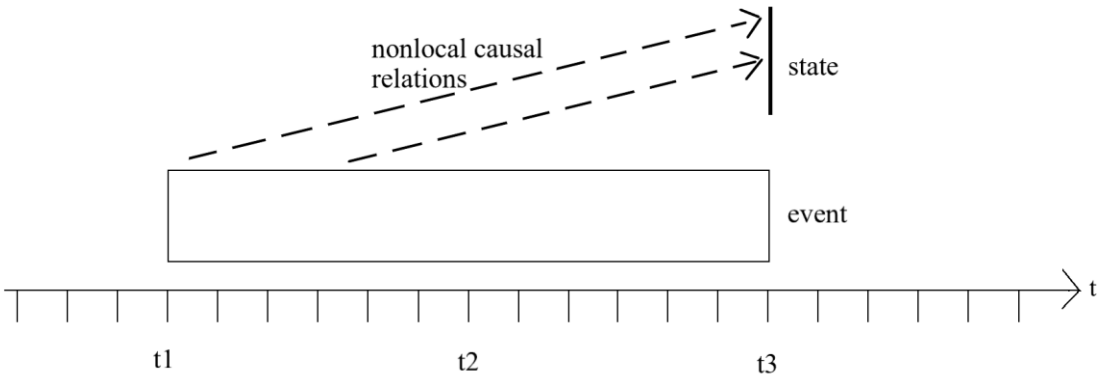


Figure 8: nonlocal causal connections

The lesson is generalizable: because a state cannot be concurrent with an event *per se*, it is impossible for that state to be non-derivatively connected to the event – connected to *all* of the event – without causal connections holding over some temporal distance. It seems, then, that it is impossible to hold that a state can be non-derivatively causally connected to an event without violating temporal locality.

It might not be obvious why this is the case. Yes, we are assuming that *B* is *non-derivatively* connected to *A.* But this does not preclude there being causal intermediaries. And so surely we can avoid temporal nonlocality by positing a chain of causal intermediaries that spans the gap between t1 and t3.

However, the problem emerges when we take a look at what these intermediaries might be. There are two options. Either they are states or they are events. Let's assume the former for now. On this view, there is a chain of causal intermediary states linking state *B* at t3 with *A* at t1. Now, if *B* is directly connected to *A*, it cannot be connected to *A* merely in virtue of being connected to *A*'s states. So the chain of intermediary states cannot connect *B* to a *state* of *A*. The chain must be connected to *A per se.* That is, it must be connected to an event. But then we face the same problem we started out with. We have an event connected to a state. And as we have seen, this violates temporal locality.

What if we assume that the causal intermediaries are events, though? We avoid the problem of temporal locality being violated at the connection with *A.* But we have the

same problem at the other end – at the connection with *B*.[9] We have to make sense of a chain of events leading to, and being causally connected to a state – *viz. B*. But again, an event cannot be causally connected to a state without violating temporal locality. Similarly, we cannot say that the chain of intermediaries comprises some combination of both events and states without running into the same issue.

Rejecting (B3) therefore forces the A-theorist to accept temporal nonlocality. I will now argue that this is problematic for the A-theorist on three counts. Firstly, it puts them at a severe dialectical disadvantage against the B-theorist. Secondly, there is no clear way for any version of the A-theory to account for the sort of causal relation they need. Thirdly, the nonlocal causal connections the A-theorist needs to posit in order to resist (B3) would commit them to an untenable model of the mechanisms involved in perception. I address these points in turn.

4.3.1 Nonlocality and dialectical considerations

In rejecting (B3), the A-theorist is committed to temporal nonlocality. This will be a bitter pill to swallow, as temporal nonlocality runs counter to what has traditionally been the philosophical and scientific orthodoxy. Indeed, temporal locality is 'almost ubiquitous in the way that scientists think about science and about what constitutes a reasonable scientific hypothesis' (Adlam 2018: 41).

Temporal locality, and *spatiotemporal* locality more generally, certainly has a long line of supporters. Take Hume, for instance, who famously claimed that 'nothing can operate in a time or place which is ever so little remov'd from those of its existence' (1740 [1992]: 75). Or Newton, who displays a more scathing attitude towards nonlocality:

> That one body may act upon another at a Distance thro' a Vacuum, without the Mediation of anything else, by and through which their Action and Force may be conveyed from one to another, is to me so great an absurdity, that I believe no Man who has in philosophical Matters a competent Faculty of thinking can ever fall into it. (Newton [in Cohen 1978: 302-3])

Certainly, nonlocality – both temporal and spatiotemporal – is highly unintuitive. In their discussion of velocity, Bigelow and Pargetter (1989) offer something of a justification.

[9] I allow for the sake of argument that the notion of a chain of causal intermediary *events* is unproblematic.

For context: they are arguing in favour of the 'flux doctrine' (that velocity is a property intrinsic to instantaneous states of objects) as opposed to the 'Ockhamist view' (that velocity is attributable to an object merely in virtue of the positions it occupies at different times).

> Consider for instance a meteor striking Mars, and consider the problem of explaining why it creates a crater of precisely the size it does. At the precise moment of impact, the meteor exerts a specific force on the surface of Mars. Why does it exert precisely that force? Because it is moving at a particular velocity. On the Ockhamist view, what this amounts to is that it exerts the force it does because it has occupied such-and-such positions at such-and-such times. In other words, the Ockhamist appeals to the positions the meteor has occupied in the past. But why should a body's *past* positions exert any force *now*? This requires the meteor to have some kind of 'memory' – what it does on Mars depends not only on its current properties, but also where it has been […] This cannot be ruled out as absurd, without further argument. Nevertheless, it is an advantage of the flux doctrine that it does not require any pseudo-memory in objects, or any time-lag in causation. (Bigelow & Pargetter 1989: 296)

Their point is that we should reject the Ockhamist view because it entails the occurrence of temporally nonlocal interactions, which in turn seems to require the operation of some sort of spooky mechanism ('pseudo-memory in objects' or suchlike). Here, then, is an argument against temporal nonlocality: All things being equal, we should avoid spooky mechanisms.

Unfortunately for the B-theorist it is far from definitive. The A-theorist can simply reply that in the current context all things aren't equal. Positing such 'spooky' mechanisms, unintuitive as doing so may be, is *less* unintuitive than what the counter-argument from experience is trying to establish – *viz.* that we don't veridically experience A-change. And so they could use the unintuitiveness of the B-theory to motivate their rejection of temporal locality. Perhaps spooky mechanisms are a small price to pay if doing so means we can preserve something so fundamental to experience as the passing of time. It seems, then, that we will have to do better than simply appealing to intuitions if (B3) is to gain any traction.

Spatiotemporal locality might instead be defended on pragmatic grounds. The idea is that if we assumed that a state of affairs has as its immediate cause some spatiotemporally distant state of affairs, it would be impossible to isolate which aspects of which states of affairs were causally relevant to any other state of affairs. The worry is that this would

undermine the whole scientific approach, as we would have no reason to suppose that the variables changed during an experiment were responsible for its outcome. By assuming spatiotemporal locality, we restrict our attention to the variables changed during the experiment, and so are able to determine the effects of those variables. As Einstein writes in a letter to Born, without the assumption of spatiotemporal locality, 'physical thinking in the familiar sense would not be possible', and it would be 'hard to see any way of formulating and testing the laws of physics' (Born 1971: 172). Indeed, given that we *do* find regularities between spatiotemporally local states of affairs, rejecting locality would seem to leave it 'completely mysterious why the effects of a cause occur at the time and place that they do' (Le Poidevin 2007a:21).

We might wonder whether these worries aren't somewhat overstated, however. Certainly, it's hard to imagine how we could even begin to go about developing a physics in a world that was properly described by laws which eschewed spatiotemporally local interactions. Of course, in such a world we would have to accept nonlocality. But we could accept nonlocality under less extreme circumstances. All nonlocality requires is that nonlocal causal interactions are possible; nonlocality could be true and it nevertheless be the case that interactions are *for the most part* local. All we would need is a theoretical framework that was able to predict under what circumstances interactions would be local and those circumstances under which interactions would be nonlocal. If we had such a framework, there would be nothing mysterious about why some interactions were local and others not. What's more, we *have* such a framework – at least, we have one that apparently predicts *spatially* nonlocal interactions: quantum mechanics.

In 1935, Einstein, Podolsky, and Rosen (EPR) proposed a thought experiment designed to show that quantum theory was, as it stood, incomplete. A simplified version of the experiment is owed to Bohm (1951), the broad strokes of which are as follows. Two entangled particles are emitted from a source in opposite directions towards spin detectors. The measurements are synchronous (or near synchronous) so as to preclude any slower-than-light signal from one measurement influencing the other, and yet there is a correlation (or rather, anti-correlation) between them: when the detectors are set to measure spin along the same axis, they will always measure opposite spins.

According to orthodox quantum mechanics, the particles' spins become determinate only once the particles interact with the detectors. But this implies that by becoming determinate, one particle's spin somehow instantaneously (or near-instantaneously)

influences the other particle's spin. Assuming that this is impossible (i.e. that spatial locality is true), EPR argued that orthodox quantum mechanics must be incomplete. The particles *must* be emitted with spin values; it's just that we can't discover what those are using the calculations currently at our disposal. So in their argument, EPR assume locality and derive the conclusion that quantum mechanics is incomplete.

However, John Bell (1964) showed that locality is in fact *inconsistent* with orthodox quantum mechanics. He devised an EPR-type test in which either detector could measure spin along a variety of axes. For each pair of particles emitted, the axis that either detector is set to measure along is randomly selected, and done so independently of the other. Given that the axes are randomly selected in this way, the results won't be perfectly anti-correlated. Sometimes both detectors will measure the same spin, sometimes opposite spins. If we assume locality, and that the particles therefore must have pre-defined spin values as ERP argued, we can calculate the probability of these results. However, as Bell established, these probabilities contradict those calculated using quantum mechanics. Simply, quantum mechanics predicts different results; it predicts results that aren't possible without nonlocality.

Experiments have since been conducted that confirm the predictions made by quantum mechanics (e.g. Aspect *et. al.* 1981; Shalm *et al.* 2015; Hensen *et al.* 2015), and it is now widely accepted that quantum mechanics demonstrates that interactions between particles in ERP-type cases are nonlocal. So it is indeed possible to formulate and test theories in physics without the assumption of locality – *spatial* locality, at least. But as it is possible with regards to spatial locality, there's no obvious reason why it shouldn't be possible with temporal locality (Adlam 2018).

Let's allow the A-theorist the point, as far as it can take them: temporal nonlocality is not ruled out. But that of course doesn't mean it's true. If the A-theorist is to reject (B3), they will first have to establish that temporal nonlocality *is* true. This presents a considerable – if not insurmountable – barrier.

The A-theorist might at this point object that the B-theorist is in the same boat. Neither temporal locality nor nonlocality have been established to be true. Temporal locality might be current scientific orthodoxy, but we have seen that it is lacking in justification. So before the B-theorist can claim (B3), they need to come up with a convincing reason for accepting temporal locality.

This is wrong, though. (B3) doesn't require the truth of temporal locality. This is because temporal locality is sufficient for the truth of (B3), but not necessary for it. (B3) claims that any perceptual awareness of extended events must be derivative if experiential states are momentary: when we are perceptually aware of an event, our experiential states must be causally connected to states of that event, rather than to the event *per se.* The truth of this claim would be guaranteed by the truth of what we might call the *causal connection principle*:

> CCP: a state can't be directly causally connected to an event; it can only be connected to a state (or states) of an event.

CCP says, in effect, that states and events can *only ever* be derivatively connected. In order to resist (B3), the A-theorist must therefore reject CCP. The truth of CCP would, in turn, be guaranteed by the truth of temporal locality.

However, CCP does not *require* the truth of temporal locality. All CCP requires is that states are only ever non-derivatively connected to other states. Nothing about this precludes the possibility of these state-state connections holding across temporal distances in the absence of intermediaries. In other words, CCP does not rule out the possibility of *nonlocal* state-state connections. So CCP is consistent with temporal nonlocality. Temporal locality is therefore sufficient but not necessary for CCP. As such, even if – and it's a big 'if' – the A-theorist could establish that temporally nonlocal causal connections are possible, they still wouldn't have shown that CCP is false.

Moreover, even if the A-theorist could somehow show that CCP is false, they still wouldn't have shown that (B3) is false. This is because CCP is sufficient but not necessary for (B3). CCP says that states and events can only be derivatively connected. If states can only ever be non-derivatively connected to states (i.e. if CCP is true), then our awareness of events must be derivative if experiences are momentary (i.e. (B3) is true). So (B3) follows from CCP. But even if (B3) is true, and no *experiential* state can ever be non-derivatively connected to any event *qua object of perception*, it doesn't follow that *no* state can ever be non-derivatively connected to *any* event. Hence, (B3) is consistent with the falsity of CCP. So even if the A-theorist can somehow show that CCP is false, they still wouldn't have shown that (B3) is false.

In resisting (B3), the A-theorist is committed against both temporal locality and CCP. However, although the truth of temporal locality and CCP would guarantee the truth of (B3), the B-theorist is committed to neither. Dialectically speaking, the A-theorist has all their work ahead of them just to break even with the B-theorist.

4.3.2 The A-theories and temporally nonlocal causal relations

Even if the truth of temporal nonlocality could somehow be established, it is not at all clear that the A-theorist could even accommodate it within their theoretic framework. I begin by examining the difficulties facing those A-theorists who subscribe to presentism. This comes in two parts: firstly, I consider the general problems posed by trans-temporal causal relations, and then the problems posed specifically by nonlocal causal relations as demanded by the rejection of (B3). I conclude that even if the presentist can answer the general challenge, their solution doesn't carry over to the nonlocal case.

I then take a look at the difficulties encountered by those who advocate the other main versions of the A-theory, namely the Growing Block and Moving Spotlight theories. I conclude that they face a dilemma. Either these theorists opt for a full-blooded or a 'ghostly' construal of non-present events. If the former, the causal relations in question can be accommodated; but doing so has unacceptable sceptical consequences. If the latter, these sceptical consequences are avoided, but then the A-theorist is lumbered with the same (or closely analogous) problems encountered by the presentist.

We begin with presentists. According to presentists, states (and events) in the past and future do not exist; all that exists exists *now*. The general challenge the presentists face is as follows.[10] Causal relations hold over time.[11] The recent stubbing of my toe is the cause of my present pain. For a relation to exist, the relata it holds between must exist; without relata there can be no relation. So for a causal relation to hold between the stubbing of my toe (X at t1) and the pain I now feel (Y at t2), both X and Y must exist. But the presentist must deny this: only the present relatum can exist – *viz.* Y. X is in the past, and so doesn't exist. As such, there can be no causal relation between X and Y. Presentists, it seems, must reject trans-temporal causal relations.

[10] My exposition of the challenge is a variation of McDaniel's (2009) formulation.
[11] It can be left open whether or not *all* causal relations hold over time.

A presentist might respond to this difficulty by providing an account of causality that does not require relations that hold between events (or states).[12] Following Baron (2012), I will call the sort of causality that has events as the relata *biff* causation. There are two ways a presentist might try to circumvent biff causation. Firstly, they might posit other sorts of causal relata. Secondly, they might construe causation in such a way as does not require the existence of relata at all.

De Clercq (2006) makes a suggestion along the former lines. The idea is that the causal relata are not taken to be events, but facts.[13] The fact that I stubbed my toe explains the fact that I am now in pain. Although one fact is a fact *about* a past event, they are both currently true; they are facts that obtain in the present. The presentist can therefore contend that these are the relata between which the causal relation holds, and in doing so avoid the problem of missing relata.

The latter approach – denying that causation requires relata – is a move available for those who espouse a counterfactual theory of causation (such as Crisp 2005; Bourne 2006; McDaniel 2009). The causal relation between the toe-stubbing and my pain would then be analysed in terms of the truth of the statement:

> If I had not stubbed my toe, I would not be in pain.

The standard Lewisian (1973) approach to counterfactuals determines the truth of this statement as follows. There are possible worlds in which my counterpart did not stub his toe. One of these possible worlds will be more similar to the actual world than any other. If, in that world, my counterpart is not in pain, the statement is true. No mention of causal relations whatsoever is needed. And if no mention of relations is needed, no mention of relata will be needed. It seems, then, that the presentist has here another way of avoiding the problem of missing relata.

Sam Baron (2012) is convinced by neither strategy. Citing the arguments of David Lewis (2004: 287), he urges that biff causation should be regarded as the most basic kind of

---

[12] Unless otherwise stated, I will assume that accounts that construe causation as a relation between events can also construe causation as a relation between states. (In fact, construing the causal relata as states rather than events avoids the problems associated with temporal nonlocality, as discussed herein.)

[13] See also Mellor (1995) Bigelow (1996) and Zimmerman (1997) who suggest we construe causal relata as belonging to a different category from events and states (as I define them). The considerations that count against DeClercq also count against these suggestions.

causation: it is what the truth of all causal statements (e.g. counterfactual statements etc.) supervene on. And so these presentist strategies avoid the problem of trans-temporal causation by ignoring the central issue.

Whether or not one finds Lewis's supervenience claim convincing for all conceivable causal statements, it certainly seems plausible when thinking about the processes involved in perception. If we are to do justice to the role played by our sense organs, we must recognise that perception is largely a mechanistic process. In the case of vision, photons with a certain wavelength *collide* with the light-receptor cells in the retina and, via certain *electro-chemical reactions*, send electrical signals up the optical nerve to the brain, and so on. This story is about biff causation. Any account that denies or ignores the interactions involved in perception will remain unsatisfactory.

How, then, might the presentist be able to account for biff causation? One suggestion would be to say that two events can be causally related so long as they have overlapping temporal parts. Or, to avoid a four-dimensionalist terminology: so long as there is a time at which the two events are present. If event X begins at t1 and ends at t3, and event Y begins at t3 and ends at t5, then the two events could be causally related at t3 (see e.g. Sider 1997: 337 fn.). This explains how causal relations hold *at* a time, but leaves it mysterious how causal relations could hold *over* time.[14] Further, it only explains how events could be *derivatively* related. This is because, assuming that the present is temporally unextended, the temporal parts of the events that could exist in the present moment would be instantaneous. Thus, the relata would be states, not events. And so the *events* would be biff-related merely in virtue of their *states* being biff-related. There would be no causal connection between events *per se.* The solution would therefore be of no help to those presentists trying to resist (B3).

It looks like the only recourse would be to drop the assumption that the present is temporally unextended. Baron (2012) considers a model of causation along these lines as a way for the presentist to account for trans-temporal causal relations. It requires the presentist accept 'thick' presentism, according to which the present is not instantaneous but has some small extension. Once they allow this, it becomes possible for them to

[14] Huemer & Kovitz (2003) argue that causal relata must be simultaneous if time is continuous. If this is right, no causal relations hold over time and so there can be no problem of trans-temporal causal relations. Far from saving the A-theorist, this would rule out the possibility of nonlocal causal relations and thus quash the A-theorist's response to (B3).

maintain that two subsequent events – or extended temporal parts of those events (which would by my definition of 'event' themselves be events) – can both exist in the present. Imagine event X exists from t1 to t3, and event Y exists from t4 to t6. If the present extends from t3 to t4, then there will be a part of X at t3 and a part of Y at t4 that the presentist can say both exist and between which a causal relation can therefore hold.

Although this model allows for trans-temporal causal relations, it comes at a heavy cost to the presentist. The idea that the present has temporal extension is highly controversial, and one that many presentists are not prepared to accept. Indeed, thick presentism is *prima facie* refuted by the following Augustinian-style line of reasoning.[15] For any temporal extension, there must be earlier and later parts. This means that, in relation to the earlier parts, the later parts are in the future; and in relation to the later parts, the earlier parts are in the past. So if any part of an interval is present, some other part of it will be in the past or in the future. Hence, if all parts of the interval are present then all parts must also be either in the past or in the future. But what is past or future cannot be present. Hence, the present cannot have temporal extension. It must instead be thought of as the durationless temporal boundary between the past and the future.

The 'thick' presentist might be tempted here to defer to quantum theory. Some quantum theorists propose that time is not continuous but granular, consisting of fundamental, indivisible units or atoms.[16] If this is true, it might support the thick presentist's claim that that the present has some small extension. Further, if these 'atoms' are indivisible in the sense that they are mereologically simple – lacking proper parts – the thick presentist will have a response to the Augustinian argument above. If the present has no temporal parts, no part can be in the past or the future in relation to another part. A present, extended chronon would be *wholly* present.

It is a matter of debate whether the idea of an extended simple makes sense, and I do not wish to venture into that arena (see e.g. Markosian 1998; Braddon-Mitchell & Miller 2006; McDaniel 2007; Sider 2007). But even *if* the idea of extended simples makes sense, and *if* time is granular, the presentist is still in a tight spot. Yes, if temporal atoms are extended simples then the presentist might be able to answer the Augustinian challenge.

---

[15] From Augustine's *Confessions* (397 – 400 AD, reprinted 1961).

[16] It is a matter of ongoing debate in quantum theory whether time is granular, and, if it is, what the size of the grains might be. Proposals range from 10^{-44} seconds (the duration of Planck time, as espoused by e.g. Rovelli 2016) to 10^{-24} seconds (the duration of a 'chronon' proposed by Caldirola (1980). The term 'chronon' was dubbed by Lévi (1927)).

But for the same reason, the thick presentist's account of trans-temporal causal relations no longer works. If the present has no temporal parts, then two events that occur at different times can no longer occupy *different parts* of the present. Only events that occur at the same time could be biff related. Biff relations holding *over* time would appear to be ruled out.

4.3.3 Presentism and the challenge posed by nonlocality

Perhaps the thick presentist would be able to argue their way out of this tight spot. Perhaps the notion of an *extended* present – even if lacking proper parts – would give them sufficient resources to tell a story about biff relations holding *over* time *within* the present. But, even if such a case could be made, there is yet a further problem for the thick presentist in trying to respond to (B3).

(B3) claims that any perceptual awareness of extended events must be derivative if experiential states are momentary. Rejecting (B3) means that the presentist must hold that the event – the object of perception – is non-derivatively causally connected to the experiential state. Assuming there is no backwards causality in operation here, the experiential state will obtain at, or some short time after, the end of the event. We may say that the object of experience, event *A*, occurs from t1 to t3 – at which point experiential state *B* obtains. As discussed above, there must be a nonlocal causal relation that holds between *A* at t1 and *B* at t3. The presentist's solution to trans-temporal biff relations considered above was to posit an extended present within which the relations can hold. To apply the same model here would means positing an extended present from t1 to t3. Why this is unacceptable becomes clear when we consider how long this interval must be.

The A-theorist's argument from experience says, in effect, that because we have experiences as of A-change, we should conclude that A-change occurs in reality. The B-theorist's counter-argument says that *even if* A-change events occur, we could only be derivatively aware of them – and these sorts of experiences cannot constitute veridical experiences as of A-change. The A-theorist says that we *can* be non-derivatively aware of events – specifically, A-change events. These are those putative A-theoretic events

such as movement.[17] So in rejecting (B3), the A-theorist thinks we can be non-derivatively aware of movement. Let's spell out what this entails for the thick presentist.

In chapter 3, if you recall, the idea of a 'borderline motion experience' was introduced. We have a borderline motion experience when we observe an object moving so slowly we are only *just* able to detect its movement. If the object was to move any slower we would not see it move. We imagined an hour hand so long that we could just about see its tip move. Let's take this to be the event the presentist claims us to be non-derivatively aware of.

Over short enough timescales, the positions of the hour hand are indiscriminable; there is a minimum duration we would need to watch it for before we could perceive any change. Following Ian Phillips (2011a), a working estimate for this duration was put at 300ms. If we were to watch the hour hand for less than 300ms, we wouldn't perceive a change. So the change we perceive occurs over 300ms or so. This, then, is the change event of which we are perceptually aware. Alarm bells should now be ringing for the presentist. We now have a timescale over which the nonlocal biff relations between the event and the mental state must hold. Event *A*, the change event in question, begins at t1 and ends at t3, which we are supposing to be 300ms later. Experiential state *B* must obtain no earlier than t3 if we are to rule out backwards causality. So there will be a nonlocal causal relation holding between t1 and t3, which is a temporal distance of 300ms. The presentist's solution to the problem of trans-temporal causal relations is to posit an extended present within which the causal relata are encompassed. But on this model, the present must have an extension of at least 300ms! Moreover, it is obscure how the presentist could account for the fact that sensitivity to motion varies from perceiver to perceiver. If one person's borderline motion experience takes 300ms, what do we make of someone else's which takes 310ms? It is difficult to see how the presentist could maintain that the present has an objective duration.

Intuitive grotesqueness aside, we can labour the point by stating the obvious fact that things happen at sub-300ms intervals. It cannot therefore be the case that the proposed extended 'present' is mereologically simple. Hence, the Augustinian refutation will apply. Further, 300ms is nowhere near the proposals found in quantum physics, to which

[17] In Chapter 3 is was suggested that what we regard as an experience of movement is in fact a combination of (at least) two perceptual mechanisms – one that detects positional change and one that detects motion *per se*, or 'pure' motion.

the presentist might be tempted to appeal. Even if the 300ms estimate is out by an order of magnitude or more, the discrepancy would still be enormous (compared with the 10^{-44} to 10^{-24} second proposals in quantum physics). Positing an extended present just isn't going to work for the presentist trying to accommodate our experience of A-change events. I conclude that if the A-theorist wants to reject (B3), they aren't going to have much success from within the presentist camp.

4.3.4 The other A-theories and the challenge posed by nonlocality

If the A-theorist is going to stand any chance of rejecting (B3), it looks like they'll have to also reject presentism. This is quite the blow for the A-theory, as presentism is often regarded as the 'purest' version of the A-theory. Other versions accept certain aspects of the B-theory's 'block universe' model. These are the 'Growing Block' and 'Moving Spotlight' theories. Both theories accept the existence of past events, but only the Moving Spotlight theory says that future events exist. So, unlike the presentist, it would appear that these A-theorists are straightforwardly able to accommodate the trans-temporal causal relations with which the presentist struggled.[18] There is a tension here, though.

According to the Growing Block and Moving Spotlight theorists, the present is somehow unique in its existence. But these A-theorists need to spell out what this means without losing grip of the claim that *non*-present events also exist. Broad (1923) did this by supposing that non-present events are different from present events *only insofar as they lack the property of 'presentness'* (call this an A-property). If this is right, the A-theorist will be able to straightforwardly accommodate trans-temporal biff relations in much the same way as the B-theorist can.

This rather weak conception of the present/non-present distinction is not a popular one among contemporary A-theorists. For one thing, it undermines one of the attractive features of the A-theory: its (apparent) ability to do justice to our common-sense temporal attitudes (Zimmerman 2011). Arthur Prior (1959) argued that if the B-theory was true it wouldn't make any sense to feel relief after some painful experience had come to an end. According to the B-theory, past pain is no less painful that present pain; both are equally real. This argument applies equally to Broad's proposal: the only difference between past

[18] For the rest of this section, when I refer to A-theorists, I am referring to non-presentists (i.e. growing block or moving spotlight theorists), unless otherwise stated.

pain and present pain is that past pain is not happening *now* – but that is the *only* difference.

Another reason for A-theorists wanting to distance themselves from Broad's conception is that it leads to the sceptical conclusion that we will most likely be mistaken any time we judge that it is the present (Bourne 2002; Braddon-Mitchell 2004; Merricks 2006). Let's say at t5 a subject judges themselves to be in the present (call this a 'NOW-judgement'). If, as Broad supposes, non-present events are just the same as present events apart from their A-theoretic properties, the NOW-judgement at t5 will seem the same to the subject whether or not t5 actually is present. So there would be no way for the subject to discern whether the NOW-judgement was true or false. Even when t5 is (briefly) present and the NOW-judgement is true, the NOW-judgement will recede into the past with the passage of time. The NOW-judgement, even if true at first, quickly becomes false – and stays false. The chances of it being true are therefore miniscule. That being so, it would be reasonable for the subject to expect their NOW-judgement to be false.

Many contemporary A-theorists avoid these undesirable consequences by stripping non-present events of properties other than mere 'presentness' (e.g. Adams 1986; Smith 1993; Williamson 1999). Zimmerman calls this the 'ghostliness strategy' (2011: 170) for obvious reasons. Although non-present events are still said to exist, the property-stripping they undergo leaves them as shadows of their former selves. Details differ between theorists, but the general idea is that although events exist at non-present times, they only occur or happen in the present. Similarly, objects involved in these non-present events exist, but non-spatially – i.e. without location, shape, size, mass and so on. This strategy enables the A-theorist to avoid the problems facing Broad's conception. I can be thankful when my toothache subsides because that dreadful event, although existent, is not occurring. And for the same reason, when I judge that an event is present, I need not worry that my judgement might be a past event.

The A-theorist needs to be wary of throwing the baby out with the bathwater, though. In stripping non-present events and objects of these problematic properties, it becomes doubtful whether they could serve as the relata of biff causality. Perception is a mechanistic process that involves the interactions of physical entities. Electrical signals propagate up the optic nerve as a result of photons *with certain physical properties* interacting with light receptor cells *with certain physical properties*, and so on. But on the current proposal, these sorts of properties only belong to the entities in question when

they are present. And so those entities will only be able to play a causal role when they are present. All non-present entities must then be discounted. This puts the Moving Spotlight and Growing Block theorists in the same position as the presentist. As such, by pursuing the ghostliness strategy, the Moving Spotlight and Growing Block theorists will find themselves facing the same problems as the presentist – and so be prone to the same criticisms.

The A-theorist seems to be facing a dilemma here, then. One the one hand they can accept a full-blooded account of non-present events and objects along the lines of Broad which equips them with the means of accommodating trans-temporal biff relations. But then they face the sceptical conclusion that our NOW-judgements are most likely false, and that many of our common-sense temporal attitudes are unjustified. On the other hand, they can avoid these pitfalls by pursuing the ghostliness strategy – but then face the difficulty of explaining how these ghostly non-present events and objects can act as relata in trans-temporal biff relations. If they can't, they will find themselves struggling against the same problems as the presentist.

4.3.5 Temporal nonlocality and the mechanisms of perception

The last criticism that can be levelled against the A-theorist trying to resist (B3) is that doing so will commit them to an untenable model of perception. We have seen that the A-theorist who rejects (B3) maintains that momentary experiential states can be non-derivatively causally connected to extended objects of experience. And we have seen that this means that they are committed to temporal nonlocality. But not only must they accept that temporally nonlocal causal connections are *possible*; they must hold that nonlocal connections play an essential role in our everyday experiences of the world. They must hold that it is *in virtue* of nonlocal connections that we perceive A-change events.

First off, there is absolutely no evidence that perception involves temporally nonlocal causal interactions of the sort that needs to be posited by the A-theorist. Any plausibility the A-theorist may have mustered in favour of temporal nonlocality has been off the back of spatial nonlocality. Quantum mechanics successfully predicts nonlocal influence in certain cases. We are justified in believing that nonlocal interactions occur because we have the weight of quantum mechanics behind us. But this theoretic framework does nothing to suggest something of a temporal analogue in the case of perception. Moreover, it does nothing to suggest a temporal analogue in the case of perception *that would suit*

the A-theorist's present needs. So the A-theorist must, in effect, be implicitly postulating some alternative theoretic framework that would do precisely the job they require of it. Physics gives us no reason to suppose such a framework is true, let alone what it might look like. But physics *does* give us quantum mechanics; and quantum mechanics does *not* deliver what the A-theorist needs.

But let's just humour the A-theorist for a minute. Ignoring all the foregoing problems, let's see what the A-theorist must really be claiming when they reject (B3). Let's say we have a momentary experiential state S in which we perceive event E. As discussed above, the A-theorist who thinks we can be non-derivatively aware of E must accept the existence of temporally nonlocal causal connections between E and S. We can talk of causal connections in terms of information transmission. When we perceive an object, we gain information about the world – information about the object. That information is transferred through the causal connection(s) holding between that object and our brain state. If we perceive E via temporally nonlocal causal connections, this means that information about E is transferred to S without first going through any intermediary mechanisms. But this makes no sense. We know how we gain perceptual information about the world: through our sense organs! In the case of vision, we know that the light receptor cells in our retinas are responsible for converting the energy of the photons that hit them into electrical impulses that propagate up the optic nerve and into the brain. If our perception of E is non-derivative, then information about E will be transferred to S through temporally nonlocal connections; and so that information will bypass our sense organs and associated neural mechanisms. On this picture, information is 'beamed' directly from the external world into our brain. If we are to account for the role of our sense organs in perception, we cannot entertain the existence of nonlocal causal connections between our experiential states and their objects. And if our experiential states cannot be causally related to external events via temporally nonlocal connections, then any perceptual awareness of those events must be non-derivative. (B3) must therefore be accepted. Our perceptual awareness of extended events must be derivative if experiential states are momentary.

4.4 Justifying (B4)

4.4.1 The problem of linguistic convention

(B4) states that our perceptual awareness of events must be accounted for within momentary experiential states. To see why this must be the case, we first need to recognise that if the success of a theory is hostage to the vagaries of linguistic convention, that theory must be rejected. This shouldn't be much of a shock. We are rightfully sceptical about theories that attach great significance to merely conventional matters. For example, astrologers would have you believe that your personality traits are determined by your zodiac sign. Because my zodiac sign is Taurus, I am supposed to have a personality similar to that of some mythological bull. It is purely conventional that we group a certain set of stars together and name it after a mythological bull. We could just as well have grouped a different set of stars together, decided it looked like a hamster, and assigned me a different set of personality traits to suit. But it was decided that *this* particular set of stars formed a constellation, and that it should be called 'Taurus'. And it is because this decision was made – because this is the convention – that astrologers think they are in a position to explain why I am stubborn, reliable, and dislike sudden changes.

I assume we are in agreement that astrology should not be taken seriously. Clearly, a theory should not rely on convention in this way; a theory should be made true by the way the world is. As I shall now argue, those theories of temporal experience that posit temporally extended experiences are also guilty of relying on convention. Instead of being made true by the way the world is they are made true by the way we *describe it*, and, as such, are guilty of 'treating language as more important than reality' (Parfit 1995: 32).

Temporal experience poses the following problem. We seem to perceive events (such as change). So it seems (as we are currently assuming to be true) that our experience represents temporally extended phenomena. As we have seen, some writers have claimed that we are able to perceive temporal phenomena events (e.g. motion) of up to 300ms in duration. However, experience is restricted to the present: what we experience we experience *now.*[19] As discussed earlier, though, the present is either a durationless boundary between the past and the future, or else it has some miniscule but indivisible

[19] That is not to say there is no time-lag between the object of experience and our perception of it. It's just that whenever an experience happens it is always temporally located – and that temporal location is what we call the present.

extension. Either way, the present would not have the extension required to accommodate our experience of change events. How, then, is it that we could be perceptually aware of events?

There are a few different ways one might respond to this puzzle. One response is to simply deny that our experience represents events, and so accept that we are therefore not perceptually aware of them.[20] This is what Dainton (2010) calls 'phenomeno-temporal antirealism'. Dainton points to Reid as an early proponent of phenomeno-temporal antirealism. Reid claimed that 'if we speak strictly and philosophically, no kind of succession can be an object of either the senses or of consciousness; because the operations of both are confined to the present point of time' (1855: 235). We must discount this response given our assumption in this chapter that premise (A1) of the argument from experience is true. We are, if you recall, allowing for the sake of argument that the A-theorist is correct in claiming that our experience represents A-change. A-change is a type of event, so if experience represents A-change then our experience must represent events. Phenomeno-temporal antirealism is not therefore on the cards as far as we are currently concerned.

We must accept for the time being that experience at least *represents* events. The responses available to us once we do so can be divided into two sorts. On the one hand we have what we may call the 'atomist' models, and on the other what we might call the 'extensional' models. Atomists contend that events are represented within instantaneous experiences. They argue that just as we should not expect an experience *of* blue to itself *be* blue, we should not expect an experience *of* an event to itself have temporal extension. Atomists are thus said to distinguish between the content and 'vehicle' of the representation. Extensionalists draw no such distinction. They maintain that events can be represented because, although events have temporal extension, so do our experiences of them. Our experiences have sufficient temporal extension to encapsulate the events being represented. The extensional theory is our target.

Because it posits temporally extended experiences, the extensional theory is hostage to linguistic convention. It was established in Chapter 2 that it is an empty question whether

[20] I take it that in order to perceive X it is necessary for one's experience to represent X. The representation of X is not sufficient for a perception of X because, as discussed earlier in the chapter, perception requires an object of experience; but a representation of X could obtain without X as an object, as in the case of hallucination.

or not experiences have temporal extension. What we normally describe as a single extended experiential event could just as accurately be described as a series of distinct momentary experiential states, metaphysically speaking. Nether description is more metaphysically accurate than the other. There is certainly no denying that it is natural to talk about experiences as extending through time. Such expressions as 'stream of consciousness' are commonplace. But as it is an empty question whether two experiential states are states of the same experience, we must accept that a different linguistic convention could just as accurately describe a succession of experiential states as distinct experiences. As such, although it is true that experiences have extension, we must accept that it is true *only because of the way we talk.* If linguistic convention was different – if we instead described the successive experiential states as distinct experiences – then it would not be true that experiences have temporal extension.

If we were to adopt this alternative convention, the extensional theory would no longer be able to explain how events are represented. The extensional theory says that events are represented within extended experiences; it says that instantaneous experiences aren't up to the job of representing events. As such, if experiences weren't extended – if we never *had* extended experiences – then the extensionalist would have to say that events cannot be represented. As discussed, however, experiences have extension merely by virtue of linguistic convention. A different linguistic convention could be adopted under which it was *not* true to say that experiences have extension. Under this alternative convention, the extensionalist theory would be unable to explain how events are represented.

As the representation of X is a necessary condition for the perception of X – it is impossible for one to perceive something that one's experience does not represent – then the extensionalist theory would be unable to account for any perceptual awareness of events under this alternative conceptual scheme. But it cannot be a matter of linguistic convention whether we perceive events. As the success of the extensionalist theory is hostage to the vagaries of linguistic convention, it should be rejected. Any perceptual awareness of events must therefore be accounted for within experiential states. This delivers (B4).

I will now consider three objections to this justification for (B4).

4.4.2 Objection 1

It is an empty question whether or not experiences have temporal extension. Because of this, we have said, experiences have temporal extension merely by virtue of linguistic convention. It was then concluded that, because of this, a theory of temporal experience cannot posit extended experiences. Surely, though, the same reasoning can be applied to those theories that posits *momentary* experiences. If it is an empty question whether or not experiences are extended then yes, we could adopt an alternative linguistic convention and eschew extended experiences. And if we adopted this linguistic convention then we could truthfully say that experiences are momentary. But although we could truthfully talk of experiences being strictly momentary, such talk would be true *merely by virtue of the linguistic convention we had adopted.* And so a theory that posits instantaneous experiences will be just as susceptible to the vagaries of linguistic convention as one that posits extended experiences. If the argument against the latter sort of theory is effective, it will be effective against the former sort. But this would mean that neither extended nor unextended experiences – i.e. *no* experiences – could be posited. This is an absurd consequence. A theory of temporal experience must be permitted to posit experiences of *some* sort. As such, the argument presented in favour of (B4) must be rejected.

This objection would succeed if the situation regarding the extensionalist and atomist theories was in the appropriate sense symmetrical. But it is not. As discussed, extensional theories cannot account for our perception of change if experiences are strictly momentary. If it were true that e1 and e2 are *not* states of the same experience – which it would be were we to adopt the alternative linguistic convention – then the extensional theory would fail. But the converse is not true for atomist theories. Atomist theories can operate whichever linguistic convention is adopted. The atomist accounts for our perceptual awareness within instantaneous experiences. In other words, they can account for our perceptual awareness with e1 alone. Similarly, they can do so with e2 alone. There is no need to appeal to an extended experience of which e1 and e2 are both states. Nevertheless, just because they don't *need* to appeal to that extended experience, they don't need to deny that it exists. As far as the atomist is concerned, e1 and e2 *could* be states of a single extended experience. It's just that the atomist's theory does not require them to posit such an experience; it would be doing no explanatory work for them. So unlike the extensional theory, the atomist theory is compatible with either linguistic convention. Hence, the atomist's theory is immune to the vagaries of linguistic convention.

4.4.3 Objection 2

It might be objected that even if we could truthfully talk about experiences as strictly momentary – if we were to adopt the alternative linguistic convention under discussion – we would nevertheless still be able to posit extended experiences. So long as we have the notion of an experiential state we can derive the notion of an experiential event, because an experiential event is simply a conjunction of successive experiential states. And so the extensional theory would work whichever linguistic convention was adopted.

What this objection misses is that the conjunctive experiential event we would end up with would be a 'fictional amalgam' (Dainton 2006: 130). This can be clearly seen by reintroducing persons *qua* subjects of experience into our discussion.

In Chapter 2 we saw how the Reductionist arguments of Derek Parfit establish not only that it *can* be an empty question whether diachronic identity holds between two subjects of experience – as Parfit argued – but that it is *always* an empty question. As it is impossible for two distinct subjects to be subjects of the same experience, it was concluded that it is always an empty question whether diachronic identity holds between two experiences.

We then considered an objection: if Reductionism is true, and facts about personal identity just consist in empirical facts, then we could cease all talk of persons and talk only in terms of the empirical facts. We could, that is, adopt an impersonal conceptual scheme which lacked the concepts of 'person' and 'subject'. This scheme would have the concept of an experience – *or more accurately, an experience** - that did not demand the existence of a subject. Once we do away with subjects of experience, however, it becomes obscure how we could conclude that it is an empty question whether diachronic identity holds between two experiences. We wouldn't be able to say it is an empty question whether two *subjects* are diachronically identical; and so we couldn't then conclude that it is an empty question whether two *experiences(*)* are diachronically identical.

To overcome this objection we took the Parfittian line of reasoning employed in the case of personal identity and set it to work in the case of experiential* identity. In short, if we can give an empirically informative account of diachronic experiential* identity – which we should, for the same reasons that support empirical informativism about personal identity – then the same conclusion follows: it is always an empty question whether

diachronic identity holds between experiences*. From then on, we restricted our discussion to experiences, and dropped the '*' from the discussion.

The objection we are currently considering claims that, even if our linguistic convention was such that experiences were essentially unextended, we could still make sense of an extended experience insofar as we could make sense of a conjunction of unextended experiences. I suggest that if one finds this objection compelling, it is because one has lost track of what an experience* is. The concept of 'experience*' would replace our concept of 'experience'. So what we would normally regard (i.e. from within our 'personal' conceptual scheme) as a single subject having a single extended experience would, from within the impersonal scheme, be regarded as a single extended experience*. Similarly, what we would normally regard as two distinct subjects having two distinct experiences would be regarded as two distinct experiences*. We can translate between the two conceptual schemes in this way.

We can assume either our normal or the impersonal conceptual scheme in responding to the objection. If we assume the normal scheme, then, as we have rehearsed just now, it is an empty question whether diachronic identity holds between subjects, and the same counts for their experiences. If it were linguistic convention that subjects – and their experiences – were temporally unextended, then it would be wrong to say we can make sense of an extended experience by simply conjoining successive experiential states. These would be the experiences of distinct subjects, and distinct subjects cannot be subjects of the same experience. As William James (1952: 147) puts it:

> Neither contemporaneity, nor proximity in space, nor similarity of quality and content are able to fuse thoughts together which are sundered by this barrier of belonging to different personal minds. The breaches between such thoughts are the most absolute breaches in nature.

If we assume the impersonal conceptual scheme, the same applies. Conjoining distinct experiences* gets us nowhere, because such a conjunction would be just as gerrymandered. Of course, we cannot appeal to the breaches between experiences 'sundered by this barrier of belonging to different personal minds'. But we can translate between the two schemes. We can appeal to the breaches of experiences between experiences belonging to different *impersonal* minds. We can suppose X and Y are experiences or experiences*. If we suppose they are experiences and decide that they

cannot be the same experience, then we would have to say that X and Y cannot be the same experience* were we to suppose they were experiences*.

4.4.4 Objection 3

It was argued in favour of (B4) that experiences have temporal extension only in virtue of linguistic convention; that it is true that experiences are extended only because of the way we talk. The charge that might be brought against this claim is that of banal triviality. Surely all truths depend on the way we talk. Take for example the truth:

T1: The Earth orbits the Sun

T1 is true, but it is only true because it is conventional to use the words 'Earth', 'orbits', and 'Sun' in such ways that make it true. If it was convention instead to use 'Earth' to refer to (what we call) the Sun, for example, T1 would not be true. But this doesn't mean that we should reject T1, or that we shouldn't let it figure in theories about celestial movements. Far from it: it is *because* of linguistic convention that we can accept T1 as true! Consider now T2:

T2: X is an extended experience.

T2 is also true by linguistic convention. But, just as linguistic convention isn't good reason to disqualify T1 from astronomy, it's not good reason to disqualify T2 from figuring in our theories about temporal experience. The extensionalist theory may depend on the linguistic convention of T2, but that is no reason to reject it.

This objection picks up on something right, but fails to recognise an important difference between astronomy and the extensionalist theory. What the objection picks up on is that there is a sense in which all true sentences are true by virtue of linguistic convention. The meaning of the words we use is a matter of convention, and so the proposition a sentence picks out will be a matter of convention. So all truths will depend on linguistic convention in this way. But once we've decided what the words in a sentence mean – once we've fixed the proposition that sentence picks out – the truth of the theory it figures in will (or, at least, *should*) depend on truthmakers 'in the world'.

Claims made in astronomy are made true by such truthmakers. And so although it is certainly the case that the truth of the theory in astronomy will depend on the way we use language, it will not be true *merely* by virtue of linguistic convention. Its truth depends,

ultimately, on the way the world is. Compare this with the extensionalist theory of temporal experience. The extensionalist theory claims that A-change is perceived within extended experiential events. Now, there is a 'thing in the world' – call it 'X' – to which the extensionalist refers to when they say 'an extended experience'. As this is the way we talk, T2 is true; X is an 'extended experience' because that is what we call it. The truth of T2 therefore depends on convention. And yes, just as this is no reason to reject T1, this is no reason to reason to reject T2. But the extensionalists' claim is more than just T2. The extensionalists claim that T2 is true, *and that it is in virtue of having these extended experiences that we are able to perceive A-change.* The point is that the extensionalists *rely* on the linguistic convention of T2 in order to *explain* how we perceive A-change. If the proposed alternative convention was adopted whereby X is described instead as a series of distinct experiences, then the extensionalist model fails. As we saw in the response to objection 1, the atomist theory is unaffected by whichever description is used. What matters for the atomist is the way the world is, not the language that it used to describe it. We can see that this is not the case with the extensionalist theory. With the extensionalist theory, the work is being done by the language used, not by the way the world is.

4.5 Summary

The A-theorist is wrong to claim that the best explanation for (A1) is that time passes in reality. Even if we allow that our experience represents A-change, this representation cannot be taken to evidence an objective phenomenon. Given the truth of Reductionism, it is an empty question whether experiences have temporal extension. Because of this, any theory that posits extended experiences in order to account for our perception of events – *viz.* the so-called 'extensionalist' theories – must be rejected. These extensionalist theories depend on the convention that we describe experiences over time as unified. If we didn't describe experiences in this way, extensionalism could not be true. Unlike the atomist theories, these extensionalist theories can appear plausible only because of linguistic convention, and so should be rejected. Any perceptual awareness of events must therefore be accounted for within momentary experiential states.

If our perceptual awareness of events must be accounted for within experiential states, then the A-theorist must accept that our awareness of events must be derivative. That is to say, any awareness we have of an event must be derived from our awareness of *states*

of that event. This is because the A-theorist is unable to hold that a state and an event could be causally connected. A causal connection of this sort would require temporally nonlocal interactions, which would mean that the A-theorist would need to posit causal connections holding between present and non-present entities. This, as we saw, puts them in a difficult position. Further, whether one accepts the A or B-theory, it is implausible to suppose that it is in virtue of these sorts of nonlocal interactions that we are able to perceive events. Our perception of events must be derivative.

Given that our perception of events is derivative, A-change cannot be an object of experience. If our awareness of events is merely derivative we would be aware of the event 'Q replacing P' in virtue of being aware of its states. We could be aware of P, and of Q, and of the ordering of those states. This would amount to an awareness of change as accepted by the B-theorist, but this does not amount to an awareness of A-change. Anything that is beyond our perceptual grasp cannot be an object of experience. So A-change could not be an object of experience if our perception of events is derivative.

Finally, because A-change cannot be an object of experience, it is wrong to suppose that an experience that represents A-change can be taken as evidence of an objective phenomenon. Even *if* change had the A-theoretic metaphysical structure *Q replacing P*, we could never veridically experience it as such. So, if we allow that A-change is represented in experience, we must accept that it is non-veridical. The best explanation for (A1) is not that time passes in reality. (A2) is refuted.

Chapter 5: A Cognitive Error Account of Temporal Experience

To recall, the A-theorist's argument from experience is:

(A1) We have experiences as of A-theoretic change.

(A2) The best explanation for (A1) is that A-theoretic change occurs in reality.

(A3) Therefore, A-theoretic change occurs in reality.

This chapter will challenge (A1) using the claim that it is an empty question whether persons persist.

Many will find the prospect that (A1) is false deeply unpalatable. There is a phenomenological dynamicity to our perception of temporal phenomena, and it is extremely intuitive to take this dynamicity as constituting an awareness of temporal passage:

> 'We perceive change and motion.' This appears metaphysically significant, because it is what 'the immediate awareness of the passage of time' amounts to. We are indirectly aware of the passage of time when we reflect on our memories, which present the world as it was, and so a contrast with how things are now. But much more immediate than this is *seeing* the second hand move around the clock, or *hearing* a succession of notes in a piece of music, or *feeling* a raindrop run down your neck. There is nothing inferential, it seems, about the perception of change and motion: it is simply given in experience. (Le Poidevin 2007b: 87)[1]

It might appear, then, that denying (A1) amounts to denying that we ever actually *see* objects move, or *hear* melodies, and so on. Barry Dainton (2010) calls those that make

[1] Le Poidevin appears to be affirming the common sense position that we have experiences as of temporal passage in virtue of perceiving change and other temporal phenomena, so it is natural to read him as affirming (A1). However, we could also read him in a deflationist light, as saying that what people *call* 'the immediate awareness of the passage of time' is *nothing more* than the perception of change – which can be given a B-theoretic description. I use the quote simply to illustrate the common sense position.

this denial *phenomeno-temporal antirealists.* However, using the term 'antirealism' in this context might lead to confusion, so I shall call the position *phenomeno-temporal error theory*, or *PT error theory* for short.[2] Those who trust their intuitions and so are committed against the PT error theory will be called *PT intuitionists.* PT intuitionism unsurprisingly has many proponents, and is often simply assumed to be true:

> …our experience of change is just as immediate as our experience of shape or colour. I take this to be an obvious truth (Dainton 2006: 114-115)

> [PT error theory is] incompatible with obvious phenomenological data (Lee 2014: 150)

> …movement and change are part of the immediate content of the most basic kinds of experience (Ishmael 2013:481)

> …we do experience constant motion: the phenomenological data leave no room for doubt (Phillips 2011a:815)

It will be useful, however, to tease apart two different claims that a PT error theorist might be thought to be making. As discussed in the previous chapter, when we say that time seems to pass what we mean is that the world seems to change A-theoretically; temporal phenomena such as motion seem to have the metaphysical structure *Q replacing P*.

The first claim, and one that I will be defending and developing in this chapter, is the claim that the content of the experience I am referring to when I say, for example, 'I am watching the second hand move around the clock' doesn't have the metaphysical structure it seems to – i.e. it doesn't represent an occurrence manifesting *Q replacing P*. Many will want to object that in saying 'I am watching the second hand move around the clock', the metaphysical structure of *Q replacing P* has already been assumed. On this view, a content with the structure *Q replacing P* is a necessary condition for the experience (as)

[2] According to Dummett (1973), a realist about domain D holds that the truth condition of a sentence about D can be *verification-transcendent*. That is, whether the sentence about D is true or not *could* be unknowable. For an antirealist, it is always knowable. What Dainton calls 'phenomeno-temporal antirealism' does not make this latter claim. On the contrary, Dainton explicitly *rejects* the view on the grounds that it is obviously false; and in doing so he is adopting a kind of antirealism about experience. As will be discussed, this is the view that the content of one's own experience is *always* knowable; there is no fact of the matter about the content of your experience over and above how it seems to you. And so the truth-conditions of sentences about experiential content are not, on their view, verification-transcendent. Those Dainton *calls* 'realists' – including himself – are, by the standard Dummettian definition, antirealists.

of motion that we are all familiar with. It is precisely this assumption that I want to challenge.

Some of the groundwork for this challenge has already been made by Christoph Hoerl (2014). He argues that the B-theorist needn't accept that our experiences of temporal phenomena constitute experiences as of passage. His suggestion is that the PT intuitionist conflates two importantly different phenomenological distinctions. As expressed by Robin Le Poidevin in the passage above, there is a clear phenomenological difference between seeing a second hand move and seeing *that* an hour hand has moved. Unlike the experience of the hour hand, our experience of the second hand has what is often called a 'dynamic' quality. The misstep as Hoerl sees it is to identify this 'dynamic' phenomenology with the phenomenology of temporal passage. The B-theorist should instead contend that the dynamic phenomenology of watching a second hand move is simply the phenomenology of *perceiving* B-theoretic 'at-at' change; whereas the non-dynamic phenomenology of seeing *that* the hour hand has moved is the phenomenology of being *inferentially* aware of B-theoretic 'at-at' change – i.e. by seeing its present position and recalling that it was in a different position previously. There is no need to appeal to experiences of different *kinds* of change to accommodate the difference in the phenomenology. I agree with Hoerl, and this chapter can be seen as building on this idea.

The second claim the PT intuitionist might be thought to be making, and one that I will *not* be defending, is the claim that the experience I am referring to when I say 'I am watching the second hand move around the clock', *whatever the metaphysical structure of its content might be*, is never in fact experienced. It would be fair enough, I think, to reject this second claim as obviously false. I can be sure that I am experiencing *this experience*, whatever the structure of its content might be. But it is not so clear, I submit, that we can disregard the first claim as obviously false in this way.

Now, some people think it is impossible to be mistaken about the content of one's own experience. This seems to be the motivation behind the PT intuitionist assertions quoted above. The thought is that, when it comes to experiential content, appearances are always as they seem (see e.g. Dainton 2013a: 390). Call this principle *SI* (for 'seems→is').

SI: what *seems* to be the content of experience *is* the content of experience.[3]

If SI is true then PT intuitionism will be true. Because change seems to have the structure of *Q replacing P*, it would be the case that the content of our change experiences really must have this structure; we really must have experiences as of A-change.

The first objective in developing a PT error theory, then, will be to establish that SI cannot be simply assumed to be true, and that accounts that deny it shouldn't be rejected merely for doing so. I will therefore take some time initially to get the reader accustomed to the idea that one can indeed make mistakes about the content of one's own experience.

Firstly, I will examine Laurie Paul's account of temporal experience. Paul is a PT intuitionist who attempts to explain temporal experience away as illusory. Her account serves nicely as a way of introducing some of Daniel Dennett's ideas, who has written extensively on the topic of subject fallibility. We see that, although Paul is a PT intuitionist, her account actually *rejects* SI in some circumstances. So the principle that would justify PT intuitionism isn't at all uncontroversial – even amongst PT intuitionists. We have good reason to reject SI, at least as an unrestricted principle (i.e. as applying to *all* experiential content).

PT intuitionism does not require the truth of SI as an unrestricted principle; it only requires it to be true with regards to the *temporal* content of experience. I continue to build the case against SI – and so undermine the justification of PT intuitionism – by considering how it fares once we restrict it to the temporal content of experience. To this end, I consider how the restricted version of SI fares in the debate between conceptualists and nonconceptualists. I show that the PT intuitionist will find themselves in a particularly weak position, coming under fire from both the conceptualists and nonconceptualists. On the other hand, in rejecting SI the PT error theorist has the advantage of being able to side with either. By exploring the options available to the PT error theorist in this way we can begin to outline what their account might look like.

Attention is then turned to putting some meat on the PT error theory's bones. The general gist is that when we report what temporal experience intuitively 'seems' like, we are prone to making certain descriptive errors. I want to show that our A-theoretic intuitions about

[3] When I use the term 'content of experience', I merely mean what the experience is (as) of. I do not wish to import more theoretical baggage than that.

time and our Non-Reductionist intuitions about the self are fundamentally connected, such that (A1) has the deep intuitive appeal it does only because we are instinctively (and inescapably) Non-Reductionists. I will argue that we should regard the intuition behind (A1) as a sort of side effect of a faulty, Non-Reductionist conceptual scheme. If we could rid ourselves of our Non-Reductionist shackles and adopt a properly Reductionist conceptual scheme, we wouldn't – indeed, *couldn't* – think that time seems to pass.

5.1 Undermining SI: Paul and Dennett

Laurie Paul is a B-theorist and a PT intuitionist. She believes that we have experiences as of temporal phenomena that constitute passage, and proposes that we explain them away as illusory. She suggests that our experience of temporal passage can be accounted for in the same way as the phi phenomenon. The phi phenomenon is apparent motion between two dots. One dot flashes up on the left side of a screen, and after disappearing is immediately followed by another dot on the right side. Watching this, it seems as though we are being presented not with two distinct dots but with a single dot moving from left to right. Even when the dots are different colours, the effect is of a single moving dot that changes colour halfway across. The traditional explanation for the phi phenomenon says that the intervening motion is somehow 'filled in' by the brain; our experience represents motion that is not there. But when we look at a footnote, we see that Paul references Daniel Dennett, who challenges this traditional explanation:

> when we have an experience as of passage, we can interpret this as an experience that is the result of the brain producing a neural state that represents inputs from earlier and later temporal stages and simply "fills in" the representation of motion [fn. not literally. It just gives the impression of being filled in. There is no "figment", as Dennett would say] (Paul 2010: 352)

According to Dennett, the representation of intervening motion is not 'filled in', nor does it need to be. After the second dot appears, the brain 'judges' that there was intervening motion, and it is because of this judgement that we report seeing the intervening motion. This judgement is what gives us the *impression* that it is filled in. Accordingly, Dennett's account 'does not include – it explicitly denies the existence of – any event which might be called the seeming-to-be-intervening-motion, on which this judgement is "based"' (Dennett 1991: 133). Whilst it *seems* that we just saw the intervening motion, we are mistaken; there was no such representation.

It might sound like this would mean Paul is a PT error theorist – after all, if temporal phenomena aren't represented, how can she claim that we have experiences as of time passing? To get a better idea of how this might work, it will be instructive to look at Dennett's treatment of another instance of 'filling in': that involved in parafoveal vision. The eye is only sensitive to detail within the small foveal range at its centre, and yet we seem to be able to see detail well beyond it. Again, the traditional explanation is that the brain 'fills in' the background of our mental image with the detail it expects to be there. This, Dennett argues, is wrong. He imagines being presented with a wall papered with the pattern of hundreds of tiny identical images of Marilyn Monroe:

> Having identified a single Marilyn, and having received no information to the effect that the other blobs are not Marilyns, [your brain] jumps to the conclusion that the rest are Marilyns […] without any further rendering of Marilyn at all.
>
> Of course, it does not seem that way to you. It seems to you as if you are actually seeing hundreds of identical Marilyns […] What is not the case, however, is that there are hundreds of identical Marilyns represented in your brain. [Your brain] just somehow represents *that* there are hundreds of identical Marilyns […] And no figment gets used up in rendering the seeming, for the seeming isn't rendered at all.' (Dennett 1991: 355)

There is a lot going on in this passage. Some have thought that Dennett is claiming a sort of perceptual illusion – that the Marilyns populate our visual field as they seem to, even though there are no corresponding *neural* representations (e.g. Pessoa *et al.* 1998; *cf.* Byrne 1998; Van Gulick 1998). However, Dennett himself has gone on to refer to the phenomenon as a 'theorist's illusion' (1998: 754). We *believe* we enjoy a mental picture in which the Marilyns are present, but the belief is false because 'the seeming isn't rendered'.

The reason for this theorist's illusion is that, according to Dennett, the brain represents *that* the Marilyns are present and detailed, without actually representing them as such. The mistake we naturally make is to assume that, because they seem to be present, they must actually be populating our mental picture. As Dennett likes to say, this is to mistake a representation of presence with the presence of a representation. It might be objected that the brain *must* be representing individual detailed Marilyns, because otherwise our visual field would be populated with ill-defined blobs – which clearly isn't the case. But Dennett has an answer: we will only see blobs if the brain represents blobs, which, again,

it doesn't do. So whilst we are safe to report how things *seem* to us, we need to be careful about what we go on to infer about the contents of our own conscious experience: [4]

> If somebody says her visual field *seems* detailed all the way out to the periphery […] there is no gainsaying her claim, but if she goes on to theorize about "the background" [of her visual field] and claims – for instance - that *there are* lots of details in this background, she becomes an entirely fallible theorist. (Dennett 2007: 263)

So there are facts about the content of our experience that are beyond our intuitive grasp. Dennett does accept that sometimes 'filling-in-beliefs may be the normal outcome of processes which include genuine filling-in' (1998: 754). But we need to be careful here. Talk of 'genuine' filling-in implies *gaps* that are filled; and if there are gaps we should ask: gaps in what? Looking back at the Marilyn passage, Dennett says the brain concludes that the parafoveal blobs are Marilyns 'without any *further* rendering' (emphasis added).[5] So it is natural to interpret Dennett here as saying that if the Marilyns *were* genuinely filled in, there *would* be 'further' rendering. He thus implies that the central Marilyns *have* been rendered.

The problem with this interpretation is that Dennett explicitly rejects the existence of a 'medium in the brain (or dualistic mind) in which subjective colours, sounds, and aromas are *rendered*' (Dennett 2015: 3). And with this he denies there are such things as '*raw feels, sensa, phenomenal qualities, intrinsic properties of conscious experiences, the qualitative content of mental states,* and, of course, *qualia'* – although he does 'agree wholeheartedly that there seem to be' (1994:129). How, then, can he talk of 'genuine' filling-in, if there is no place where any rendering happens? Well, Dennett does allow that our normal ways of talking about consciousness will generally make sense so long as they are 'immersed in a bath of metaphor' (Dennett 2002: 14). So he is presumably talking about 'genuine filling-in' in strictly *metaphorical* terms. The difference between 'genuine filling-in' and 'non-genuine filling-in' is then to be cashed out in terms of brain-level (rather than experiential) representations. In both cases the brain will represent *that* there is 'filling in' – hence the subject's 'filling in' beliefs. But only genuine 'filling in' is

[4] Dennett is not alone in thinking so. It is commonly held that psychology experiments investigating change blindness and other similar phenomena demonstrate that we intuitively suppose visual experience to be much richer than it is (e.g. O'Regan 1992; Blackmore *et al.* 1995; Simons & Levin 1997; Rensink 2000; Ballard 2002; Noë 2007; Schwitzgebel 2008).

[5] This phrase is not a one-off; he repeats it elsewhere (e.g. 1992: 47; also 2015: 6).

accompanied by the representation *of* whatever the filling-in belief refers to.[6] Importantly, though, in neither case is there any experience-level 'rendering'.

We can now see that there are actually two ways in which the subject is prone to error in the phi illusion on Dennett's account. Firstly, the ostensible 'filling in' seems to be a case of 'genuine filling-in', but isn't. The brain merely represents *that* there was intervening motion. Secondly, in reporting to see the dot move across the screen, it is assumed that experience has the sorts of phenomenal properties Dennett explicitly rejects. This second claim is a considerably bolder one, for it means that a subject is in error *any* time they refer to the qualitative content of their mental states. It seems to the subject their experience has a qualitative content, and it is precisely this content to which a subject's phenomenological reports refer, but really *there is no such content.*[7]

This latter view is by no means uncontroversial. However, the fact that there is no agreement whether or not it is right means that we *cannot* assume subject infallibility. Further, even critics of this stronger claim can agree with the weaker claim that phenomenological reports can be – are often grossly – inaccurate (e.g. Schwitzgebel 2007 and Velmans 2007). So although Paul is a PT intuitionist, upon close inspection her account provides the PT error theorist with the leverage they need to get their foot in the door. SI as a general principle is controversial even amongst PT intuitionists, and, moreover, evidence in psychology (for example, that the visual field doesn't represent as much detail as we intuitively suppose) shows that there are cases in which it *can't* be true.

Perhaps the best thing for the PT intuitionist would be to restrict SI to temporal experience. Experience, it could be argued, is *essentially* dynamic, such that an experience that did not represent time passing would be no experience at all. So perhaps there is a special reason to suppose that SI more plausibly holds in the case of *temporal* experience. Call this restricted version of the principle *TSI* (for 'temporal seems→is').

---

[6] Because in non-genuine filling in cases there is still a representation *that*… , it is impossible for the subject to tell the difference between genuine and non-genuine cases on the basis of experience alone. Of course, that does not mean there is no fact of the matter; the concern is epistemological, not metaphysical.

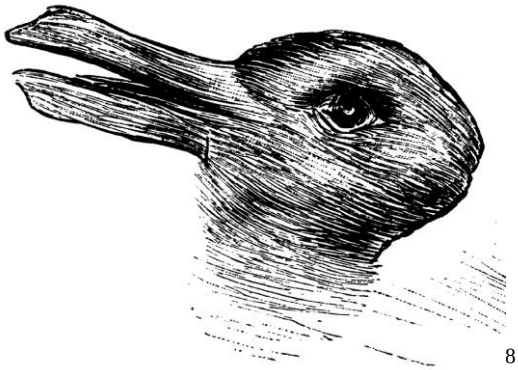
[7] That is not to say Dennett rejects the existence of consciousness *per se*. He acknowledges that 'there has to be something – some property K – that distinguishes conscious events from nonconscious events' (1994: 58). He just doesn't accept that consciousness has the sorts of properties we naturally ascribe to it. How Dennett thinks we *should* talk about the 'experience' is not my concern here.

TSI: what *seems* to be the temporal content of experience *is* the temporal content of experience.

5.2 Undermining TSI: the (non)conceptualism debate

Restricting SI to the temporal content of experience might overcome some initial qualms. But when we consider what it would entail in the debate between the conceptualists and nonconceptualists we can see that it still puts the PT intuitionist in a difficult situation. As will be shown, any PT intuitionist who assumes TSI will find themselves coming under attack from both sides of the debate. The PT error theorist on the other hand, is free to side with either. We start off with some definitions.

Conceptualists hold that all mental content is a function of the concepts possessed by the subject (e.g. McDowell 1994). For example, when you look at a table, your perceptual content is determined by your concept of 'table' and the concepts it encapsulates – such as 'legs', 'surface' and so on. It then follows that if you didn't possess those concepts the phenomenology would be different. It is important to make explicit that, assuming perceptual content is conceptual, it will be determined not only (assuming perceptual content is conceptual) by the concepts a subject *possesses*; it must also matter which of those possessed concepts are 'experientially active'. To illustrate, you may possess the concept of, say, 'fulfilment', but that plays no part in shaping your experience of the table. Rather, it must be the concept of 'table' that is doing the work. The duck-rabbit optical illusion demonstrates the point nicely:



[8]

[8] Unattributed drawing from the 23 October 1892 issue of German magazine 'Fliegende Blatter'.

You possess the concepts of 'duck' and 'rabbit', but the phenomenology only presents one at a time. As such, concept possession *alone* cannot be what determines the content. When you are presented with duck-phenomenology, your concept of 'duck' must somehow be involved in a way that your concept of 'rabbit' is not (and *vice versa*). This is what I mean by 'experientially active'. The duck-rabbit illusion is remarkable insofar as, because you are able to shift the phenomenology at will, you are to a great extent in control of which concept is experientially active. In most cases, however – like when you look at a table – this is not the case. Experiential content is then standardly defined as conceptual iff the subject's undergoing the experience entails that they possess the concepts that canonically characterise it (Crane 1992: 382). 'Canonical characterisation' is theory-relative, meaning that a PT error theorist and a PT intuitionist will disagree over what concepts characterise the temporal content of experience.

Conversely, nonconceptualists hold that *at least some* mental content is nonconceptual. Nonconceptualists typically point to the content of perceptual experience as an example of nonconceptual content, but can allow that other sorts of mental content, such as the content of propositional attitudes, are conceptual.[9] Perceptual content is then standardly defined as nonconceptual iff the subject need not possess the concepts that canonically characterise it.

5.2.1 PT intuitionism and (non)conceptual content

Call the temporal content of experience C. PT intuitionists think that our intuitive characterisation of C is correct (*viz.* in terms of A-change concepts)[10]. Further, by assuming TSI they must think that our intuitive characterisation *cannot be mistaken*.[11] As such, there can be no situation in which a subject characterising C lacks the appropriate concepts. So by PT intuitionism,

[9] I will be restricting my discussion to the content of perceptual experience unless stated otherwise. I use 'experience', 'perception', and 'perceptual experience' interchangeably.
[10] By 'A-change concepts' I mean those that refer to temporal phenomena with an assumed metaphysical structure of *Q replacing P*.
[11] Unless otherwise stated I will be assuming from here on that TSI is used by the PT intuitionist to justify (A1). I also assume that the language used in the characterisations accurately reflects the subject's beliefs about what they are experiencing.

(i) Characterising C entails possession of the concepts that correctly characterise C.

This gets the PT intuitionist very close to treating C as conceptual. Recall that C is conceptual iff:

(ii) Being subject to C entails possession of the concepts that correctly characterise C.

By (ii), a subject of C is guaranteed to possess the appropriate concepts irrespective of any characterisations they happen to make. So *a fortiori* when the subject turns their attention to characterising C they will have the correct concepts at their disposal with which to do so. It therefore suits the PT intuitionist nicely to accept (ii) and treat C as conceptual. But can they treat C as *non*conceptual? Well, if they were to reject (ii), theirs would be the view that a subject undergoing C is guaranteed to possess the correct concepts in virtue of *characterising* C. This is a bizarre view (what could it be about the act of *characterising* that brings this guarantee?) and not what the PT intuitionist wants to say. For the PT intuitionist, a subject is able to correctly characterise C simply in virtue of *undergoing* it. Being able to correctly characterise entails possession of the correct concepts, so a subject undergoing C would be guaranteed to possess the correct concepts in virtue of *experiencing* C. In other words, if you are subject to C you will possess the correct concepts. The PT intuitionist therefore holds (ii), and so is committed treating C as conceptual.

This does not mean the PT intuitionist is committed to conceptualism, though. Firstly, conceptualists think that *all* content is conceptual, but the PT intuitionist needn't go so far. As long as they think some other content isn't conceptual, they will be a nonconceptualist. Secondly, the conceptualist's reasons for treating C as conceptual are in stark contrast to the PT intuitionist's. The PT intuitionist argues that (ii) is true on the grounds that:

(iii) A subject possesses the concepts that characterise the content by virtue of experiencing it

whereas, as we have seen, the conceptualist holds that (ii) is true on the grounds that:

(iv) A subject experiences a particular content in virtue of their experientially active concepts.

So the PT intuitionist gives explanatory priority to the content, whereas the conceptualist gives explanatory priority to the concepts. Accepting (iv) and giving priority to the concepts is not an option for the PT intuitionist. (iv) entails that subjects with relevantly different experientially active concepts will experience time differently. However, it is implicit in the PT intuitionist's position that there is only one sort of phenomenology to be characterised; there is something it is like to experience time, and that something is characterised by A-change concepts. If we allow the assumption that we all experience time in more-or-less the same way then, by (iv), the PT intuitionist must assume we all possess relevantly similar experientially active concepts.[12] This seems reasonable enough.

But what if the PT error theorist had their way? The error theorist contends that we need to abandon A-theoretic concepts and find a better way of making sense of C. But because the PT intuitionist thinks that our A-theoretic concepts are experientially active, they will also see a danger in abandoning those concepts: the risk of experiential *de*activation of those concepts, and a resulting phenomenological shift. The PT intuitionist cannot allow this to happen, because then our A-change concepts would no longer correctly characterise the experience.

Imagine looking at the duck-rabbit image and only ever seeing the duck, completely unaware that it can ever look like a rabbit. You insist that the visual content is characterised by 'duck' – much as the PT intuitionist insists that C is characterised by A-change concepts. But imagine that one day someone comes along and says to you "we can do better than 'duck' – it makes more sense to think of it as a rabbit". So you look at the image with the concept of 'rabbit' in mind, your experientially active concept changes to 'rabbit', and lo and behold you start seeing the rabbit instead of the duck. And, importantly, what used to be the correct characterisation ('duck') is no longer correct.

Clearly, the PT intuitionist cannot afford for this to happen; if it did they wouldn't be able to maintain that A-change concepts characterise the phenomenology. As such, they would need to insist that alternative concepts would *not* be able to characterise the

[12] It is not out of the question that there could be individuals who are neurologically different in some way that means they do not experience time in the same way as the rest of us. The discussion is therefore restricted to 'normal' neurological functioning.

phenomenology in the way that 'duck' and 'rabbit' both can in the duck-rabbit example. Unlike in the duck-rabbit case there is only one way of correctly characterising the phenomenology – and this is how we currently characterise it. But then their position begins to look rather desperate. Even if we cannot *currently* imagine what it might be like, with *our* conceptual scheme, who knows what we would be inclined to say about C were we to assimilate a different conceptual scheme? There is the possibility – an empirical matter – that with a different conceptual scheme we might not characterise C as we do now. This might happen naturally over time, or, more fantastically, as the result of some sort of Big Brother-esque censorship program. This is not important. What is important is that the possibility would force the PT intuitionist to admit that C might be better characterised without A-change concepts.

The PT intuitionist therefore needs to block the error theorist's move to abandon A-change concepts, which they can only do – as they have been doing – by insisting on our current characterisation as the correct one. But as we have just seen, if our current characterisation is correct, this might just be because a different way of conceptualising the experience hasn't yet been successfully assimilated. As such, it doesn't look like the PT intuitionist who holds (iv) can dismiss PT error theory without begging the question. The PT intuitionist shouldn't therefore accept (iv). And if they reject (iv), they set themselves against the conceptualist.[13]

To summarise so far, then. The PT intuitionist holds (ii) and so treats C as conceptual. But this presents them with something of a dilemma. On the one hand, they could reject

[13] One notable conceptualist is McDowell. McDowell claims that, because perceptual content is able to feature in our judgements, it must be conceptual. It doesn't matter whether we have the language to characterise it; what matters is that it *could* be referred to – in thought at least. Giving it a name is simply to 'give linguistic expression to a concept' that we already possess (2008: 753). All perceptual content can be referred to in thought, so all such content is conceptual. Whether or not McDowell is right about this, it doesn't seem to be of any help to the PT intuitionist. McDowell is still telling a story where concepts explain content, and as we saw the PT intuitionist must reject this idea. Besides, it is not entirely clear that McDowell is presenting a position that opposes nonconceptualism as defined above rather than simply defining conceptual content differently. If any content that can be referred to is conceptual content, then any content that can be characterised is conceptual content. But nonconceptual content was defined earlier as content that could be experienced by a subject *without that subject possessing the concepts that characterise it*. This allows that the content *could* be characterised, were that subject to possess the appropriate concepts. But if it could be characterised, then according to McDowell it is conceptual. It looks like the disagreement might be merely over the definition of 'nonconceptual'. Indeed, it has been argued that McDowell's conceptualism and the sort of nonconceptualism that populates much of the literature are in fact compatible (Mazijk 2015).

nonconceptualism by claiming that *all* mental content is conceptual. But if they do so they will set themselves against the conceptualists by disagreeing over whether the concepts or content has explanatory priority. As we saw, conceptualists affirm (iv) – that a subject experiences a particular content in virtue of their experientially active concepts – whereas the PT intuitionist seems forced to deny (iv). On the other hand, the PT intuitionist could reject conceptualism by allowing that *some* mental content is nonconceptual. The problem with this move is that the majority of the arguments advanced by nonconceptualists – indeed, the literature in general – claim that *perceptual* content is nonconceptual (e.g. Peacock 1992; Davies 1997; Dretske 2008; Evans 1982; Crane 1988(a) & (b), 1992; Bermúdez 1994, 2009); it is the content of propositional attitudes that tends to be regarded as conceptual. So if they reject nonconceptualism, the PT intuitionist will also set themselves against the nonconceptualist. The PT intuitionist is therefore in a very weak position. By relying on TSI to justify their position they will, whichever way they turn, find themselves alone and under attack from both sides.

Enough has now been said to discredit the PT intuitionist's reliance on TSI to justify their position. It is wrong to dismiss PT error theory as obviously false; we must take seriously the possibility that we are mistaken about what our temporal experiences represent. Perhaps they don't represent occurrences with the metaphysical structure of *Q replacing P*.

5.2.2 PT error theory and (non)conceptual content

We can begin to get an idea of what a PT error theory might look like by considering the PT error theorist's options in the (non)conceptualism debate. The PT error theorist thinks we don't have experiences as of A-change. Theirs is the position that our intuitive characterisation of temporal experience is somehow faulty. The position is essentially a negative one; it states how we *shouldn't* think about experience, but doesn't specify how we should. Whether or not we possess the concepts to correctly characterise temporal experience is left open. Perhaps we can adequately characterise temporal experience using concepts we already have at our disposal, or perhaps we can't; perhaps it *can't* be characterised.

If the PT error theorist thinks C cannot be characterised, then they will be a nonconceptualist. If they think it can be characterised but we lack the appropriate concepts, then *a fortiori* being the subject of C does not entail possession of the concepts

that correctly characterise C, and the content is nonconceptual. So the PT error theorist can straightforwardly be a nonconceptualist.

If we are equipped with the correct concepts, C could be regarded as either conceptual or nonconceptual. If the PT error theorist thinks our possession of those concepts is entailed by having the experience, they will regard C as conceptual. This is the route a PT error theorist must take if they are to be a conceptualist. If on the other hand they think *we just happen* to possess the concepts, they will regard C as nonconceptual. Either way, it will be the PT error theorist's view that despite our possession of the correct concepts, we nevertheless identify the wrong ones when characterising C. We have already seen that the PT error theorist can unproblematically be a nonconceptualist. So what are the prospects for them if they are to take the conceptualist route?

You would be forgiven for thinking the prospects don't look good. How could the PT error theorist claim that experiencing C entails possession of the correct concepts – and stronger still, that those concepts are experientially active when we experience C – whilst *also* claiming that we wrongly identify those concepts as A-change concepts? As it turns out, this is quite a natural claim for the PT error theorist to make as a B-theorist.

The B-theorist contends that change is nothing more than temporal variation, whereas the A-theorist insists on the *Q replacing P* structure. If P is replaced by Q, then Q obtains after P. Temporal variation is necessary for A-change, in the sense that A-change could not occur without temporal variation obtaining. And so without the concept of temporal variation we can't have the concept of A-change. The concept of temporal variation is, if you like, 'built into' the concept of A-change.

Because the PT intuitionist thinks that C represents A-change, they must also think that C represents temporal variation. And if the PT intuitionist thinks that A-change concepts correctly characterise C, they must also think that the concept of temporal variation is necessary for the correct characterisation of C. Further, if the concept of A-change is experientially active, then, as the concept of temporal variation is built into the concept of A-change, so the concept of temporal variation must be experientially active.

The PT error theorist – if they are to take the conceptualist route – can accept this story in part. They can agree that the concept of temporal variation is experientially active, and that it is necessary for the correct characterisation of C. But they can contend that this

concept is sufficient, and that the intuitionist's error is in supposing that the A-theoretic component is also necessary. So the PT error theorist can, as a B-theorist, unproblematically regard C as conceptual. And so, provided they think all other mental content is conceptual, can be a conceptualist.

I do not want to argue that a PT error theorist should be a conceptualist or that they should be a nonconceptualist. But we at least now have an idea what they can claim, whichever way that debate might be resolved. Unlike those who assume SI, the PT error theorist has the flexibility to side with either.

5.3 The error

So far I have been trying to lend plausibility to the idea that our intuitive characterisation of temporal experience is wrong. But it will no doubt still be obscure *how* we could make a mistake about the content of our experience. That is what I want to focus on next. It will help to begin with an example of the sort of error I will be proposing for our PT error theorist account.

5.3.1 A cognitive error

Consider what it's like when you push a finger down onto a hard object like a table. Nowadays we know that the vast majority of the mass of an object is concentrated in the nuclei of the atoms that comprise it. And we know that these nuclei are separated by empty space – so much of it that the composition of objects is 99.9% empty space. If we were to ask someone in the eighteenth century if the table felt like it was mostly empty space, they would say 'no'. They would say the table feels 'solid' – by which they would mean the table feels 'saturated' or 'full' of mass. Call the property of being 'full of mass' *solidity**. Even though today we know that tables are not solid*, we might even be tempted to agree that they nevertheless *feel* solid*, devoid of the empty space we know they largely consist of. But it would wrong to conclude that our, or anyone's experience for that matter, represents solidity*; it would be a mistake to characterise our experience as being as of solidity* - even as an illusion.

We understand and characterise the content of a non-veridical experience on the basis of what the experience would be of, were it veridical. We know what it is like to have a veridical experience of a red ball, so we know what an experience *as of* a red ball is like

– whether it be veridical or non-veridical. An experience *of* X is in this sense logically prior to an experience *as of* X. We could say that non-veridical experiences are parasitic on veridical experiences in this way.[14] So to find out what an experience *as of* solidity* would be like, we need to ask what a veridical experience *of* solidity* would be like. What would a veridical experience of solidity* be like, then? Well, if a volume the size of a table was saturated with mass – if the atoms were compressed such that there was no empty space between the nuclei and electrons – then there would be all manner of forces in play, and touching it would be nothing like touching a table. It can't therefore be right to say that we ever have experiences as of solidity*, even illusory ones. Instead, our deeply-held pre-scientific intuitions about mass and what it should feel like lead us astray when we characterise our experiences.

This example shows us something important about the relationships between phenomenology, belief, and the content of experience. 'Phenomenology' is traditionally put in terms of 'seemings': the phenomenology of X is how X *seems* to the subject. But as the solidity* example shows, a subject's phenomenological reports – what they would say an experience *seems* like – is infected by their beliefs and intuitions about the way the world is. Someone from the eighteenth century believes that tables are solid*, and they have no reason to believe their experience isn't representing the world accurately, so they would say the table seems solid*. But this shows that what a subject's experience is *actually representing* does not always align with how it seems to them. A subject's beliefs and intuitions about the way the world is can lead them astray in characterising the content of their experience. The error is not perceptual; it is, if you like, cognitive. This is the basis of the PT error theory I will be proposing. Just as faulty intuitions lead us astray in characterising the content of experience when we touch hard objects, so our faulty intuitions lead us astray when we characterise temporal experience. More specifically I propose that what leads us astray when we think about our experience of time is the way we think about ourselves as subjects of experience. The way we intuitively think about

[14] One potential objection with this principle – that veridical experiences are logically prior to non-veridical experiences – might make use of the waterfall illusion. In the waterfall illusion we have an experience as of a scene which appears to be both moving and not moving. So the content of the experience is contradictory. But it is surely impossible to have a *veridical* experience with a contradictory content, so it cannot be the case that non-veridical experiences are parasitic on veridical ones. We can respond to this objection by utilising the distinction that was drawn in chapter 3 between the perception of 'pure motion' and the perception of positional change. The waterfall illusion does not have a contradictory content so long as we distinguish between these two representations. The waterfall illusion represents 'pure motion' without representing positional change (see Mellor 1988 for a similar argument).

ourselves is essentially Non-Reductionist. If we could free ourselves from our Non-Reductionist shackles, it would make no sense to suppose that time seems to pass. If we could assimilate Reductionism into our belief system, we wouldn't mischaracterise temporal experience.

The outline of the account is as follows. The belief that we have experiences as of A-change necessarily presupposes the concept of personal identity. Without the concept of personal identity – that the same subject can exist at two different times – it would make no sense to suppose that a subject could experience P at t1 *and* Q at t2.[15] That is, it would make no sense to suppose that one could experience *Q replacing P*. And if it would make no sense to suppose that one could experience *Q replacing P*, it would make no sense to suppose that one could have an experience *as of Q replacing P*; if it is impossible to conceptualise an experience *of* X, it is impossible to conceptualise an experience *as of* X.

Because Parfittian Reductionism is true, a conceptual scheme which lacked the concept of personal identity would be just as metaphysically accurate as ours. This Reductionist scheme should be preferred, as were we equipped with it we would not fall into the Non-Reductionist trap of thinking that our persistence is metaphysically significant. And, further, we would not fall into another Non-Reductionist trap of thinking that time seems to pass.

Parallels can be drawn between my proposal and Simon Prosser's (2016). Prosser also denies that we experience time as passing, since he considers the notion of temporal passage to be unintelligible. If experience were to represent time passing, it would therefore have a 'necessarily false' content. But, Prosser urges, experience cannot have a necessarily false content; so it cannot represent time passing. Prosser is, as such, a PT error theorist.

However, Prosser is not always read as a PT error theorist. Hoerl (2014) for example appears to have him down as a PT intuitionist (i.e. someone who thinks temporal passage *is* represented in experience). The reason for this misclassification might, I suggest, be due in part to Prosser's account of why time *seems* to pass. He thinks that the putative 'dynamic' phenomenology of temporal experience can be explained by our experience representing objects as enduring. He invokes the Kantian idea (1787 [1996]: 257) that

[15] I use 'person', 'self', and 'subject' interchangeably.

change requires identity: 'only the permanent (i.e. substance) undergoes change […] this permanent makes possible the presentation of the transition from one state to another'. It is natural to read 'the transition from one state to another' as describing A-theoretic change. Now, Prosser thinks 'permanence' *is* represented – insofar as we have experiences as of enduring objects. So, since he holds that permanence *is* represented in experience, and he invokes the claim that this 'permanence' is required for the presentation of (A-theoretic) change, we might reasonably expect him to also hold that A-change *is* (re)presented. The fact that he doesn't think so might be thought to create something of a tension in his account; and one that is perhaps evidenced by Hoerl's misclassification of his position.

What I propose shares the general idea that change requires identity, but differs in several respects. Prosser proposes, in short, that time seems to pass because:

> Experience (mis)represents objects (and ourselves *qua* subjects of experience) as enduring.

Whereas I propose that time seems to pass because:

> We conceive of ourselves *qua* subjects of experience as persisting.

There are a few key differences. Firstly, my account locates the error at (and restricts it to) the level of conceptualisation, rather than experience. Because of this, the account sits squarely within the providence of the PT error theorist, and avoids any tension that might be thought to be present in Prosser's account. It says that the error is cognitive, rather than perceptual.

One immediate objection here would be that the distinction between conceptualisation and experience might not be so cut-and-dried: how we conceptualise the world bleeds into the way we experience it. This is the basic idea behind conceptualism (but a nonconceptualist would presumably be willing to accept this claim up to a point). To respond to this objection, we may reconsider the solidity* example. The solidity* example shows that we *can* distinguish between perceptual and cognitive errors. As was argued, we cannot regard the experience we have when we touch solid objects as a perceptual illusion. Our experience *cannot* be representing solidity*; the mistake can only be one of characterisation. Further, it is not even clear that this commits us against the

claim that the way we experience the world depends on our concepts. All we need to claim is that we are able to distinguish between perceptual and cognitive *errors*. And we don't even need to claim that there is always a neat distinction between perceptual and cognitive errors; it will suffice to show that *in some cases* there is. And this is precisely what the solidity* example establishes.

Secondly, my account focuses exclusively on the identity of subjects, and ignores objects. This is because, for it to make sense that one could experience A-change, it is necessary and sufficient that subject identity is preserved through the change; but it is neither necessary nor sufficient that object identity is. To explain: If you experience P (and not Q) and I experience Q (and not P), clearly neither of us has an experience of Q replacing P – even if the set of objects in P and Q are the same. So subject identity is necessary, but object identity is not sufficient, for the experience of A-change. On the other hand, P and Q could be states of affairs with distinct objects. When we awake suddenly from a vivid dream, for example, the objects that were represented are replaced by an entirely new set. So we can make sense of experiencing A-change without object identity, provided the same subject experiences both sets of objects. Hence, subject identity is sufficient, but object identity is not necessary.

Thirdly, Prosser's account focuses on a particular kind of persistence (*viz.* endurance), whereas mine is concerned with persistence *per se.* Prosser's account relies on the falsity of endurance theory to explain temporal phenomenology. The problem with this approach is that endurance theory is naturally aligned with the A-theory, so Prosser's account won't do anything to convince many A-theorists that temporal passage is not represented in experience. Further, it's not clear that B-theorists are forced to reject endurance theory, so his solution won't necessarily appeal to all B-theorists either. The account I propose, on the other hand, engages both sides of the debate. Everyone should recognise that the way we naturally conceive of ourselves as subjects of experience is flawed. Whether you think that persons persist by perduring, enduring, or exduring, you must, if you accept Reductionism (which you should), accept that we could equally well describe persons as non-persisting.

5.3.2 Empty questions and alternative conceptual schemes

As was argued in chapter 2, it is an empty question whether a person persists through empirical change. On the assumption that over any interval (no matter how brief) there

will be some empirical change (no matter how trivial), it follows that it is an empty question whether persons persist *per se.* As such, what we ordinary describe as a single persisting person could just as accurately be described as a series of distinct persons. Indeed, it could be described as a series of (infinitely many, if time is dense) momentary persons. My disagreement with someone who denies that I am the same person who began writing this sentence – or even that I am the same person from one moment to the next – would be merely verbal. It would not be like the substantive disagreement we would have with someone who thinks the Sun orbits the Earth. It would be like disagreeing with someone about whether Jurassic Park is an 'old' film. Or, more accurately, it would be like disagreeing with someone who contends that *no* films are 'old'. Any failure on our part to accept this is due to our inability to shake off the Non-Reductionist shackles.

If a single persisting person could just as accurately be described as a series of distinct, momentary persons, then a conceptual scheme which lacked the concept of diachronic personal identity would be just as metaphysically accurate as ours. Whereas we use 'I' to refer to our past and future selves, an alternative scheme could use 'they', and 'we'. Under this 'we' scheme, you would regard your past and future selves as temporal ancestors and descendants rather than as *you*. If someone were to ask me 'what are you up to at the moment?', I would reply something like 'we're just working on the thesis'. Thinking in this way is of course difficult to imagine. And perhaps for some it is not pleasant to imagine. Brueckner puts it nicely:

> Maybe one [scheme] is more comforting than another, or maybe one has some other desirable consequence. But if the question really is empty, then no [scheme] is better than another in respect of *truth*, or of *conformity to the facts*. (Brueckner 1993, p. 15, emphasis in original) [16]

We can speculate about the practical benefits of an 'I' scheme over a 'we' scheme. Prosser (2016: 183) suggests that a perceptual system that represents objects as enduring would be more computationally economic than one that represents them as perduring. The same kind of consideration holds for a cognitive system that conceptualises the self as a persisting entity rather than (infinitely many) momentary ones. Further, thinking of my

[16] To be accurate, Brueckner's quote is not one in which he is comparing the two conceptual schemes as I am. Rather, he is comparing the two ways of describing an outcome when it is an empty question. We could *describe* a situation as involving one person, or we cod describe it as involving two distinct persons. But neither *description* is better than the other in terms of 'conformity to the facts'. What Brueckner says for descriptions goes equally well for the two schemes I am comparing.

temporal descendants as *me* means that I have a unique sort of future-directed concern that has clear evolutionary benefits. I am avoid painful situations because *I* don't want to feel pain; I sacrifice immediate gratification so that *I* may reap the long-term rewards. That is not to say I don't care whether others feel pain, or that I never act in ways that benefit others over myself. But the drive to do so is different when I think of the future person as *me*.

A 'we' scheme would have its advantages, though. In thinking of our temporal relatives as 'others' we would be liberated from Non-Reductionism. We wouldn't fall into the trap of thinking that our persistence is metaphysically significant. And, as the concept of personal identity is necessary for the belief that A-change is experienced, we wouldn't fall into the trap of thinking that time seems to pass either. Without the concept of personal identity it would be impossible to believe that I could be the subject of two states of affairs obtaining at two different times. It would make no sense to suppose that I am experiencing *Q replacing P*.

This is the general thrust of the account. There is still more to be explained, but it might be helpful to do so in response to some objections one might have.

5.3.3 Objections

The first objection is that my account will do nothing to convince PT intuitionists that they are wrong. My claim is that time would not seem to pass under a 'we' scheme; that it would make no sense to suppose that we have experiences as of A-change. Even if this claim is true, it does not show that (A1) – that we have experiences as of temporal passage – is false. The question of whether or not we have experiences as of A-change is left open. The PT intuitionist thinks we do have experiences as of A-change, and so they would regard my account merely as showing that, *were* we to adopt a 'we' scheme, we would be prevented from recognising the fact that we do indeed have that kind of experience.

Further, because of this, they would be unwilling to accept my claim that a 'we' scheme would be just as metaphysically accurate as our 'I' scheme. Our 'I' scheme makes it possible for us to recognise the existence of a certain kind of experience – *viz.* those that represent A-change. If these sorts of experience would be absent from a 'we' scheme, then the 'I' and 'we' accounts are not equally legitimate ways of conceptualising and describing the world. The 'I' scheme should therefore be preferred. And so the proposed PT error theory could never even get off the ground.

What this objection brings to the fore is the need to distinguish between two parallel 'I' schemes: a Reductionist one and Non-Reductionist one. As argued, Reductionists should accept that persons persist through empirical change merely as a matter of linguistic convention. It is convention to *say* that persons persist, but this persistence is not 'metaphysically significant'. This runs counter to our deeply held intuitions about personal persistence. I cannot help but feel that my persistence *is* metaphysically significant. Try as I may, I cannot shake the instinct that my diachronic identity is more than a matter of linguistic convention. When I take action to avoid pain, it is not because I can *call* that future person who would otherwise have suffered 'me'; it's because they *are* me - or so my intuitions tell me. As a Reductionist, I believe these intuitions to be faulty, but they are so ingrained in the way I think that it is seemingly impossible to override them. I am intellectually Reductionist, but instinctively Non-Reductionist. The first-person concepts I cannot help but employ invoke the metaphysical significance that I, as a Reductionist, regard as irrational.

We can, however, recognise the possibility of a different conceptual scheme that contained first person concepts, but in which those concepts did not invoke the metaphysical significance that ours do. Under such a scheme, 'I' would be directly equivalent to 'we' in the 'we' scheme. As it stands, our concept 'I' is *not* equivalent to the 'we' scheme's 'we' – precisely because of the metaphysical significance first person concepts invoke. So although the two schemes would use the same first person *words* – 'I', 'me' etc. – the concepts they contained would be importantly different. The concepts in our 'I' scheme are Non-Reductionist in that they invoke, or assume, the metaphysical significance entailed by Non-Reductionism, but we can recognise the possibility of a Reductionist 'I' scheme in which they don't.

Now that we have distinguished between Reductionist and Non-Reductionist 'I' schemes, we can see that my earlier claim that an 'I' scheme and a 'we' scheme would be equally legitimate – as equally accurate metaphysically speaking – is in need of clarification. A 'we' scheme would be as metaphysically accurate as a *Reductionist* 'I' scheme. But, because our *Non-Reductionist* 'I' scheme attaches unwarranted metaphysical significance to first person concepts, a 'we' scheme would actually be *more* metaphysically accurate than our 'I' scheme.

This answers the objection above. Yes, our 'I' scheme enables the belief that we have experiences as of A-change. But our 'I' scheme is Non-Reductionist, and as such it is

defective. A Reductionist 'I' scheme would correct this defect. To get an understanding of the implications adopting a Reductionist 'I' we can look to those of a 'we' scheme, since a 'we' scheme would be as metaphysically accurate as a *Reductionist* 'I' scheme. Under a 'we' scheme it would make no sense to suppose that we have experiences as of A-change; and so under a Reductionist 'I' scheme it would make no sense either. It is therefore only because we possess a faulty Non-Reductionist 'I' scheme that (A1) seems plausible. Could we correct our conceptual scheme, and fully realise the superior way of conceptualising reality that a Reductionist scheme would enable, we would not say that time seems to pass.

The second objection one might have concerns the proposed account's focus on subjects of experience, rather than the experiences themselves. And in doing so, the objection would run, I have made the assumption that experiences are individuated by subjects. Only if it is impossible for a single experience to be experienced by distinct subjects would it make no sense that two distinct subjects could have an experience of *Q replacing P*. And so the 'we' scheme would preclude the thought 'I am experiencing A-change' only if the concept of 'experience' is such that experiences are individuated by subjects. Even if experiences are individuated by subjects according to our conceptual scheme, that doesn't mean the same would be true of a 'we' scheme. Perhaps under the 'we' scheme it would be natural to suppose that a series of distinct subjects *could* share the same experience. And if that were the case it would still make sense to suppose that A-change could be experienced. So it cannot be claimed that a 'we' scheme wouldn't yield the intuition that we have experiences as of A-change.

What this objection highlights is that the operative concept is not *personal* identity *per se*, but rather *experiential* identity. Personal identity is relevant only insofar as our conceptual scheme individuates experiences by subjects. But, if we are Reductionists, we do not need the concept of 'subject' in our conceptual scheme. If personal identity just consists in empirical facts, all we need in our scheme are the concepts that refer to those lower-level facts. We could do away with the concept of 'subject' altogether (Parfit 1999). We could adopt an *impersonal* conceptual scheme, and this scheme would be just as accurate as the 'I' or 'we' scheme.

Our concept of 'experience' comes attached to the concept of 'subject'. An experience can't *be* an experience unless it is experienced by a subject. This is a matter of definition. An impersonal conception scheme however would have the concept of 'experience*',

which does not demand the existence of a subject.[17] 'I am experiencing A-change' would then translate to something like 'this is an experience of A-change'. In order for this to make sense, we would need the concept of experiential identity; that the same experience could be an experience of P at t1 and of Q at t2. And from here we can see that the same line of argument relating to personal identity will apply to experiential identity. The same epistemological arguments that are effective against Non-Reductionism about personal identity are also effective against Non-Reductionism about experiential identity. Consciousness is not a non-physical substance; if we are to say something informative about what experiential identity consists in, we must say something *empirically* informative. That being the case, questions of experiential identity will be empty. As such, an alternative conceptual scheme which lacked the concept of experiential identity would be just as accurate as one that didn't. And, under such a scheme, it would make no sense to suppose that an experience could be *both* an experience of P at t1 *and* Q at t2. The experience of P at t1 could not be the *same* experience as that of Q at t2. It would therefore make no sense to suppose that an experience could be an experience of A-change; and so the thought that an experience could be an experience *as of* A-change would be unthinkable.

5.4 Summary

PT error theory is highly unintuitive – so unintuitive, it is often assumed to be false. But the principle that would justify such an assumption – that what *seems* to be the content of experience *is* the content of experience – needs to be treated with suspicion. As we have seen, this principle is controversial even within the PT intuitionist camp. Further, evidence from psychology – for example, regarding the amount of detail represented in visual experience – shows that in some cases it *is* possible to be mistaken about the content of one's experience.

Maybe the PT intuitionist would concede that the principle is not true for *all* cases, but maintain that, at least in the case of temporal experience, it is. The problem with this manoeuvre is that someone can only make it at the expense of facing a dilemma in which they fit in with neither the conceptualists nor the nonconceptualists. The PT error theorist

[17] I assume that we are happy to talk of experiences without subjects, and drop the '*' from the discussion.

on the other hand has the freedom to be a conceptualist or a nonconceptualist. PT intuitionism cannot be assumed true; PT error theory deserves a seat at the table.

The B-theorist may therefore reject the A-theorist's motivating intuition that experience represents time passing. Provided, that is, they can explain how it is we come to make such a mistake. It has been proposed that this mistake results from our Non-Reductionist conception of the self. We naturally suppose that our persistence is metaphysically significant, that questions of identity are not empty. This conception of the self is built into our conceptual scheme in the form of first person pronouns. Because Reductionism is true, however, questions of identity are empty. As such, a 'we' scheme would be just as accurate as an 'I' scheme. Further, once we distinguish between Reductionist and Non-Reductionist 'I' schemes, we can see that our current scheme is a *Non-Reductionist* 'I' scheme; and that a (Reductionist) 'we' scheme would therefore be metaphysically superior. The intuition that I experience A-change – *Q replacing P* – necessarily presupposes that I am the same subject that experiences both states of affairs. Without the concept of personal identity (and experiential identity), it would make no sense to suppose that I could experience A-change. Such an experience would be impossible to conceptualise. And so it would be impossible to conceptualise an experience *as of* A-change. We would be faced with a certain phenomenology, but we could not characterise this phenomenology as time passing.

Conclusion

The A-theorist's argument from experience is simple and intuitive, and captures the main motivating intuition behind the common sense belief that time passes:

(A1) We have experiences as of A-theoretic change,

(A2) the best explanation for (A1) is that A-theoretic change occurs in reality;

(A3) so, A-theoretic change occurs in reality.

I have argued that neither of its premises should be accepted by anyone who holds a Reductionist view of personal identity.

The first job was to define Reductionism, and show why Non-Reductionism should be rejected. This was the focus of Chapter 1. Parfit's own definitions weren't particularly helpful, even in his attempt at clarification in his later work. We saw that his definitions would end up counting some Cartesian Views as Reductionist, despite the fact he explicitly classifies them (and argues against them) as varieties of Non-Reductionism. The diagnosis of the problem was that Parfit defined Reductionism in terms of the grounding relation only, but without any mention of the entities between which the relations hold. Having taxonomised the various possible positions, it was clear that a satisfactory definition needed to account for both.

Since Parfit's own attempts at defining Reductionism weren't particularly helpful, it was necessary to 'reverse engineer' a definition. This was done by examining the arguments against Non-Reductionism, the idea being that those views not targeted would have a good claim to qualify as Reductionist. The epistemological arguments against the Cartesian Views were straightforward. In short, since we have no evidence of separately existing entities we shouldn't believe they exist. Arguments against the other view Parfit explicitly classifies as Non-Reductionist – the Further Fact View – were not so straightforward. Since the proponent of the Further Fact View is a *primitivist* about

personal identity, they were able to resist the arguments of Shoemaker and Parfit. Other arguments against the primitivist were supplied by the Criterialists. These arguments focused on the disastrous epistemological consequences of denying the existence of necessary and sufficient conditions for persistence. Although there is logical room for a primitivist who accepts necessary and sufficient conditions for persistence, it was concluded that such a view was implausible for a number of reasons. Finally, I argued that the considerations that count against the primitivist's Further Fact View also count against the 'brute persistence views': the views which say that personal identity consists in the brute persistence of an empirical entity. The only view left untouched by the criticisms was the view that personal identity consists in empirical relations holding between empirical entities. This gave us our definition of Reductionism.

In Chapter 2 we examined Parfit's Combined Spectrum thought experiment. Parfit's target was the intuition that there is *always* a *deep* difference between identity and non-identity. The argument he provided purported to show that there can be no sharp borderline or cutoff anywhere on the spectrum (where on one side we had a clear case of identity, and on the other a clear case of non-identity). He took this to show that identity can be indeterminate; and he claimed that when a question of identity has an indeterminate answer, that question is empty. By supplementing Parfit's discussion with that of Ted Sider's notion of nonsubstantivity, it was shown that Parfit's claims were in need of clarification.

Firstly, his claim that there could be no sharp borderline on the spectrum makes sense only if we read Parfit as saying that there can be no *substantive* sharp borderline. That is to say, there can be no borderline on the spectrum where on one side it is substantively true that the person persists, and on the other side it is substantively true that they don't. That would require there to be a metaphysically privileged entity that the term 'same person' picked out, and that could make such claims true. Only the Non-Reductionist can claim there to be such an entity. There *can*, however, be many *nonsubstantive* borderlines on the spectrum (where we would be merely *describing* a case as involving identity or non-identity).

Secondly is the claim that identity can be indeterminate. It was shown that it cannot be *substantively* true that identity is indeterminate, since that would require a metaphysically privileged *vague* entity – which, again, only the Non-Reductionist can posit. But identity will be indeterminate in the sense that it will be *neither substantively true nor*

substantively false that identity holds. Questions of identity can also have indeterminate answers in the sense that, since it is an empty question whether identity holds, we can legitimately *describe* identity as *neither holding nor not holding.* Both these sorts of indeterminacy result from the fact that questions of identity are, given Reductionism, empty, or nonsubstantive.

I then argued for a stronger conclusion that Parfit's: given the truth of Reductionism, it is *always* an empty question whether a person persists through empirical change. Parfit's restriction is unjustified. This claim was then extended to experiences: on the assumption that it is impossible for two persons to be subjects of the same experience, this means that it is always an empty question whether experiential identity holds. Experiences have temporal extension, but only because of the way we talk. It would be no less metaphysically accurate to describe experiences as strictly momentary states.

In Chapter 3 this 'Neo-Parfittian' claim about experiential identity was defended against the arguments of Ian Phillips and Matthew Soteriou. Both argued that there are cases where we can only make sense of momentary experiences as *part* of extended experiential wholes. If true, the question of whether an experience has temporal extension would not be empty, since experiences would be *fundamentally extended*. Coincidentally, Soteriou and Phillips both argued that their preferred 'holistic' conception of experience was uniquely placed to make sense of certain aspects of motion perception.

I began with Soteriou's argument. Soteriou argued that a puzzle was brought about by the conjunction of the following two assumptions: firstly, that a subject's awareness of an observed event has the same temporal extension as the observed event, and secondly, that when a subject is in a given state over an interval, this is determined by the fact that they are in that state at each moment during the interval. These two assumptions were puzzling because a subject who is aware of an object's movement from location L1 at t1 to L10 at t10 can be said to be in a *state* of awareness that obtains from t1 to t10; but if they are in a state of awareness then we must be able to attribute them that same state – being aware of the object's motion from L1 to L10 – at each moment during that interval. But at t1, the object has yet to reach L10; so we *can't* attribute them such a state of awareness! Soteriou claimed that only by accepting the holistic conception can we make sense of this.

I countered by arguing that the puzzle can only be generated by presupposing the 'atomistic' conception – the conception of experience that Soteriou is arguing against – to be false. If the atomist accepts that the proffered mental state obtains they will have to deny the second of Soteriou's assumptions; and if they accept the second assumption they will have to deny that the subject is in the proffered state. Either way, there is no puzzle. Soteriou therefore failed to show why the atomistic conception of experience should be rejected. Further, the solution he offered required us to reject a representationalist theory of perception and commit ourselves to a relationalist theory instead. The atomist solution made no such demands, and so is the preferable option.

Next we looked at Phillips's argument, which made use of a puzzle raised by Delia Fara. The puzzle resulted from the two obvious facts that, firstly, we experience constant motion, and secondly, some positional changes are too small to be perceived. Phillips's solution had two commitments at its heart: that experience has a 'temporal field', and that the content of this field is purely determinable over short enough sub-intervals. Phillips claimed that these commitments can only be accepted by the holist.

My response to Phillips was to point out that his notion of a temporal field as a sort of window through which we can directly perceive extended events had unacceptable neurological consequences, and ran together two importantly distinct neural mechanisms: one that computes positional change, and one that computes 'pure' motion. By separating these mechanisms out, a much simpler, atomistic solution became available. All we need to make sense of the two troublesome facts that Delia Fara raised is the claim that a subject enjoys non-illusory motion phenomenology when and only when an observed object is in a sufficiently different position from a sufficiently recent position it was observed in. Neither Soteriou nor Phillips succeeded in undermining the atomistic conception of experience. Far from it; the puzzled they presented can be dealt with far more effectively by the atomist than the holist.

Having defended the Neo-Parfittian claims about personal and experiential identity, they were put to work against the A-theorist's argument from experience in Chapters 4 and 5. Chapter 4 focused on (A2) of the argument from experience: the claim that the best explanation for our experiences as of passage is that time really passes. I argued that, even if we allow that our experience represents A-change, this representation would not provide evidence of an objective phenomenon.

If it is an empty question whether experiences have temporal extension, any theory that posits extended experiences in order to explain how we perceive events will depend for its success on linguistic convention. Hence, these theories should be rejected. Theories that posit momentary experiences, I argued, do not suffer from this problem. Hence, our perceptual awareness of events must be accounted for within momentary experiential states.

Given that only those theories that posit momentary experiential states are tenable, our awareness of events must be 'derivative'; when we are aware of an event, this awareness must be derived from our awareness of instantaneous states during that event. Several reasons were given for this. For one thing, if our awareness of events was *non*-derivative, information about the world would have to be 'beamed' straight into our brains, bypassing our sense organs and the associated causal mechanisms. Further, we would have to accept temporal nonlocal interactions. Although controversial, this might not be such a problem for a B-theorist; but it was shown to pose deep problems for the A-theorist. Our awareness of events cannot be non-derivative.

It was then argued that, since our perception of events must be derivative, A-theoretic change cannot be an object of experience. A-change, I claimed, is an event with the metaphysical structure of 'state Q replacing state P'. If our awareness of events is derivative, then our awareness of A-change would be derived from our awareness of state P, and of state Q, and of the ordering of those states. But we would not be aware of the *replacement.* The obtaining of states P and Q in an order is something the B-theorist accepts, so at most this awareness would be an awareness of B-theoretic change. Anything that is beyond our perceptual grasp cannot be an object of experience, and, as such, A-change could not be an object of experience. If A-change cannot be an object of experience, then we could never have a veridical experience of A-change. But if we could never veridically experience A-change, then, if we are to assume that A-change is at least represented in experience, we must accept it as a *misrepresentation*. As such, this representation cannot be taken to evidence an objective phenomenon. (A2) was thereby refuted.

The final chapter of the thesis addressed (A1) of the argument from experience: that we have experiences as of A-change. In response to this premise I proposed a cognitive error account of temporal experience.

To many, it will seem *just obvious* that we have experiences as of A-change. And so the prospects for an error theory did not look good. Efforts were made to soften these intuitions. We began by considering a general principle that would justify the common sense view that such an error theory would be obviously wrong: that what seems to be the content of experience *is* the content of experience. If this principle were true, we could never be mistaken about the content of our experience. I firstly discredited this principle by showing that it is controversial even amongst those who think we do have experiences as of A-change. Further, evidence in psychology shows that we *do* make mistakes about what is represented in experience – such as when we uncritically suppose our visual field is far more detailed than it actually is – and that these mistakes can be ubiquitous. Further, if an account like Dennett's is true, then we are *radically* mistaken about the content of our experience.

It would be a mistake to suppose that we are infallible judges of our own experiential content. The prospects for a restricted version of the principle – that we are at least infallible judges of the *temporal* content of our experience – were considered. By situating this principle within the debate between conceptualists and nonconceptualists, its proponent was shown to face a dilemma. Conceptualists think that, if we are subject to a particular experience, we will possess the concepts to correctly characterise that experience. Since the proponent of the restricted infallibility principle thinks that it is impossible to be wrong about the temporal content of experience, they must hold that we possess the correct concepts to characterise it. Hence, they think the temporal content of experience is conceptual. This is where they face the dilemma: they could either be a nonconceptualist (if they think that at least *some* mental content is nonconceptual), or a conceptualist (if they think *all* mental content is conceptual). If the former, they will find themselves alienated from the rest of the nonconceptualists, since nonconceptualists argue that experiential content *isn't* conceptual; but if the latter, they will alienate themselves from the rest of the conceptualists by disagreeing over whether the concepts of content has explanatory priority. Whichever way, they will find themselves facing stiff opposition from both sides of the debate. But, by rejecting the infallibility principle, we are free to side with either.

This opened the way for a cognitive error theory. I suggested that the mistake we make when we characterise temporal experience can be thought of as the same kind of mistake we make when we characterise hard objects as feeling as though they are 'saturated with

mass'. When we say that objects do not *seem* to be mostly empty space, we should not say our experience is illusory; rather, we simply have a false belief bout what the representation is (as) of. The question then of *why* we have the false belief in the case of temporal identity was addressed.

It was proposed that, since it is an empty question whether a person persists, an alternative conceptual scheme which didn't contain the concept of diachronic personal identity – a 'we' scheme – would be just as metaphysically accurate as an 'I' scheme that did. Moreover, we can distinguish between two sorts of 'I' scheme: our Non-Reductionist one (in which persistence come attached with the notion of metaphysical significance) and a Reductionist one (in which it doesn't). I argued that a 'we' scheme would be preferable to our 'I' scheme. Under such a scheme I would use 'we' to refer to a part of my temporally extended self that included my present state; and 'they' when referring to a past or future self. I would consider my past and future selves as *distinct* subjects of experience. I would still be able to recognise and value the special kinds of relations that would hold between me and these temporal counterparts – in much the same way that I do with those I love, share ancestry with, and have memories in common with. But, in much the same way, I would regard them as 'others'.

The consequences of fully assimilating such a conceptual scheme would be far reaching, and throw into question the justifications for many of our common practices. For example, the notion of responsibility would be upended: how can I be praised or blamed for someone else's act? Similarly, acts of future-directed self-interest might be reasonably viewed as a sort of nepotism.[1] Parfit argued that identity was not what matters in survival, and used this claim to construct an ethical framework around. The position I have been arguing for would no doubt require an extension of his ideas, but much work would need to be done in order to determine exactly what this would amount to. For now, however, the consequence we are interested in is the impact this way of thinking would have on our beliefs about time and our experience of it.

The A-theoretic notion of temporal passage, we have been supposing, is of a type of temporally extended event with the metaphysical structure of 'state Q at t2 replacing state P at t1 as the presently obtaining state'. In order for it to make sense that a I could experience A-change, therefore, I must suppose that I experience both states. But since

[1] That is not to say these acts would be *unjustifiable*; but if we were to continue practicing them, we would presumably need to alter the ways in which we justified them.

the states obtain at different times, this would require the concept of diachronic personal identity. Without the concept of personal identity, we would not intuit that both states were experienced by the *same* subject. I could not intuit that I – a single momentary subject – experience A-change. Since the content of a non-veridical experience of X is parasitic on the content of a veridical experience of X, I could not therefore intuit that I have experiences *as of* A-change. The intuition expressed in (A1) of the argument from experience should therefore be considered as a sort of side-effect, or by-product, of a faulty conceptual scheme.

The Neo-Parfittian account presented does, however, leave some questions unanswered - most notably, if the intuition that time seems to pass is a side-effect of a faulty Non-Reductionist conceptual scheme, how then *should* we characterise temporal experience? What would we say were we to adopt a properly Reductionist conceptual scheme? This will be an important question for future research.

A fruitful place to start might, I suspect, be Le Poidevin's (2007b: 95) suggestion that we project A-theoretic properties onto a B-theoretic world, in much the same way that projectivists think we project secondary qualities onto the world.[2] The projectivists' basic idea is that secondary qualities – such as colour – do not belong to the objects that we ascribe them to. Rather, the *primary* (e.g. light reflective) qualities of the object cause us to have an experience of colour, which we then project back onto the object.

Now, we couldn't give this exact account in the case of A-change if the phenomeno-temporal error theory is true. A projectivist account would say that certain primary qualities in the world cause us to have experiences as of A-change, which we then project back onto the world. But the central claim of the account I have proposed is that we *don't* have experiences as of A-change in the first place. There might nevertheless be something worth pursuing in the idea that temporal passage is a secondary quality.

One way of distinguishing between primary and secondary qualities in the Lockean tradition is brought out in the following passage:

> [T]o grasp the concept of red [a secondary quality] it is necessary to know what it is for something to look red, since the latter constitutes the satisfaction condition for an object's being red. In contrast, to grasp what it is for something

[2] See e.g. Boghossian & Velleman 1989 for a projectivist account of colour. Le Poidevin (2007b: 96) defends projectivism about A-theoretic properties against an argument by Gale (1968).

> to be square [a primary quality] it is not constitutively necessary to know how square things look or feel. (McGinn 1983: 8)

Since we don't need to experience primary qualities in order to know what they are, we might say that concepts for primary qualities can be grasped 'purely intellectually'. We can understand the concept of 'square' so long as we have the concept of 'line', 'angle', 'point' (and so on), and understand how they relate to one another. We might say that 'square' is conceptually structured in this sense. On the other hand, we can't grasp the concept 'red' without experiencing redness. We might say that secondary qualities are essentially sensible properties.

Intuitively, hard objects feel solid* - they seem to feel 'saturated with mass'. This concept can be grasped purely intellectually. We don't need to know what a solid* object would look or feel like to be able to understand what 'solid*' means. Our grasp of 'solidity*' comes from our grasp of 'mass', 'saturation', etc., and how these concepts are related in instances of 'solidity*'; we understand what solidity* is because we have a grasp of its conceptual structure. Hence, solidity* is a primary quality. However, I have argued that the experience does not represent solidity*. One suggestion would be that the experience doesn't represent a primary quality at all. Instead, we mistake the representation of a secondary quality for a representation of a primary quality. We are subject to a representation of an essentially sensible property, but we misattribute this representation with a conceptual structure.

Such an account would take a great deal of developing. But if it could be made to work in the case of solidity*, there is no obvious reason it couldn't work in the case of temporal experience. The idea would be that the phenomenology we associate with temporal passage – what is often called a 'dynamic' phenomenology – would be that of a representation of a secondary quality. We then misattribute this essentially sensory representation with the conceptual structure of A-theoretic change. What is responsible for this misattribution is, I have argued, our Non-Reductionist conceptual scheme.

I hope to have provided a compelling case against the A-theorist's argument from experience. Neither premise should be accepted by anyone who isn't prepared to face the disastrous epistemological consequences of Non-Reductionism. Intuitively, our persistence is metaphysically significant. And, intuitively, the belief that time passes is justified by our experience of it. These intuitions are connected. Since our experience is

not metaphysically significant, we cannot trust the intuition that time passes, or even that we have experiences that represent it as such. It might be impossible to fully embrace Reductionism. Nevertheless, so long as we recognise its truth, we must try to resist our A-theoretic tendencies. We don't have experiences as of temporal passage, nor if we did could it be explained by time passing.

Bibliography

Adams, R. 1986. Time and Thisness. In P. French, T. Uehling & H. Wettstein (eds.) *Midwest Studies in Philosophy: Volume 11*. Minneapolis: University of Minnesota. 315-329

Adlam, E. 2018. Spooky Action at a Temporal Distance. *Entropy.* 20(1): 41-61

Alter, T. & Rachels, S. 2005. Nothing Matters in Survival. *The Journal of Ethics.* 9 (3-4): 311-330

Augustine, St. 1961. *Confessions.* R.S. Pinecoffin (ed.). Harmondsworth: Penguin

Aspect, A., Grangier, P. & Roger, G. 1981. Experimental tests of realistic local theories via Bell's theorem. *Phys. Rev. Lett.* 47(7): 460—463

Ballard, D. 2002. Our perception of the world has to be an illusion. *Journal of Consciousness Studies.* 9(5-6): 54-71

Baron, S. 2012. Presentism and Causation Revisited. *Philosophical Papers.* 41(1): 1-21

Baron, S., Cusbert, J., Farr, M., Kon, M. & Miller, K. Temporal Experience, Temporal Passage and the Cognitive Sciences. *Philosophy Compass.* 10(8): 560-571

Behrendt, K. 2003. The New Neo-Kantian and Reductionist Debate. *Pacific Philosophical Quarterly.* 84(4): 331-350

Bell, J.S. 1966. On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys*. 38: 447 [Accessed online March 2 2019] https://doi.org/10.1103/RevModPhys.38.447

Belshaw, C. 2011. Animals, Identity and Persistence. *Australasian Journal of Philosophy*. 89(3): 401-419

Bermúdez, J. L. 1994. Peacock's Argument Against the Autonomy of Nonconceptual Content. *Mind and Language.* 9(4): 402-418

Bermúdez, J. L. 2009. The Distinction Between Conceptual and Nonconceptual Content. In B. McLaughlin, A. Beckermann & S. Walter (eds.) *The Oxford Handbook of Philosophy of Mind.* Oxford: Oxford University Press. 457-473

Bigelow, J. 1996. Presentism and properties. *Philosophical perspectives.* 13: 35-52

Bigelow, J. & Pargetter, R. 1989. Vectors and Change. *British journal for the philosophy of science*. 40: 289-306

Blackburn, S. 1997. Has Kant Refuted Parfit? In J. Dancy (ed.) *Reading Parfit*. Oxford: Blackwell. 180-201

Blackmore, S., Brelstaff, G., Nelson, K. & Troscianko, T. 1995. Is the Richness of our Visual World an Illusion? *Perception.* 24: 1075-81

Boghossian, P. & Vellenam, D. 1989. Colours as a Secondary Quality. *Mind.* 98: 81-103

Bohm, D. 1951. *Quantum Theory*. New York: Prentice-Hall

Born, M. 1971. *The Born-Einstein Letters.* New York: Walker & co.

Bourne, C. 2002. When an I? A tense time for some tense theorists? *Australasian Journal of Philosophy.* 80: 359-371

Bourne, C. 2006. *A Future for Presentism*. Oxford: Oxford University Press

Braddon-Mitchell, D. 2004. How do we know it is no now? *Analysis*. 64: 199-203

Braddon-Mitchell, D. and Miller, K. 2006. The Physics of Extended Simples. *Analysis*. 66: 222-6

Broad, C.D. 1923. *Scientific Thought.* Oxford: Oxford University Press

Brueckner, A. 1993. Parfit on What Matters. *Philosophical Studies.* 70(1): 1-22

Byrne, A. 1998. Dennett versus Gibson. *Behavioural and Brain Sciences.* 21: 751-752

Cassam, Q. 1989. Kant and Reductionism. Review of Metaphysics, 43(1): 72-106

Caldirola, P. 1980. The introduction of the chronon in the electron theory and a charged lepton mass formula. *Lettere al Nuovo Cimento*. 27(8): 225–228

Carter, W. R. 1999. Will I be a Dead Person? *Philosophy and Phenomenological Research.* 59: 167 - 172

Chuard, P. 2011. Temporal Experiences and Their Parts. *Philosopher's Imprint*. 11(11): 1-28

Chisholm, R. 1967. Identity Through Possible Worlds: Some Questions. *Nous.* 1(1): 1-8

Cohen, I. 1978. *Isaac Newton's papers and letters on natural philosophy.* Cambridge, MA: Harvard University press

Craig, W.L. 2000. *The Tensed Theory of Time: A Critical Examination.* Dordecht: Kluwer Academic Publishers

Crane, T. 1988(a). The Waterfall Illusion. *Analysis.* 48(3): 142-147

Crane, T. 1988(b). Concepts in Perception. *Analysis.* 48(3): 150-153

Crane, T. 1992. The Nonconceptual Content of Experience. In T. Crane (ed.) *The Contents of Experience.* Cambridge University Press. 380-401

Crisp, T. 2005. Presentism and "Cross-Time" Relations. *American Philosophical Quarterly.* 42(1): 5-17

Curzer, H. 1991. An Ambiguity in Parfit's Theory of Personal Identity. *Ratio,* 4(1): 16-24

Dainton, B. & Bayne, T. 2005. Consciousness as a guide to personal persistence. *Australasian Journal of Philosophy.* 83(4): 549-71

Dainton, B. 2006. *Stream of Consciousness.* Oxford: Routledge

Dainton, B. 2008. *The Phenomenal Self.* Oxford: Oxford University Press

Dainton, B. 2010. Temporal Consciousness. In E. Zalta (ed.) *Stanford Encyclopedia of Philosophy.* [Accessed online 5 October 2016] https://plato.stanford.edu/archives/win2018/entries/consciousness-temporal/

Dainton, B. 2013a. The Perception of Time. In H. Dyke, & A. Bardon (eds.) *A Companion to the Philosophy of Time.* Sussex: Wiley-Blackwell. 389-409

Dainton, B. 2013b. Time, Passage, and Immediate Experience. In C. Callender (ed.) *The Oxford Handbook of Philosophy of Time*. Oxford: Oxford University Press. 382-419

Dancy, J. 1985. *Introduction to Contemporary Epistemology.* Oxford, Blackwell

Davies, P.C.W. 1995. *About Time: Einstein's Unfinished Revolution.* Harmondsworth: Penguin

Davies, M. 1997. Externalism and Experience. In N. Block, O. Flanagam, & G. Güzeldere (eds.) *The Nature of Consciousness.* London: MIT Press. 309-327

De Clercq, R. 2006. Presentism and the Problem of Cross-Temporal Relations. *Philosophy and Phenomenological Research.* 72(2): 386-402

Dennett, D. 1991. *Consciousness Explained.* USA: Penguin

Dennett, D. 1993. The Message is: there is no *Me*dium. *Philosophy and Phenomenological Research.* 53(4): 919-931

Dennett, D. 1994. Real Consciousness. In A. Revonsuo & M. Kampinnen (eds.) *Consciousness in Philosophy and Cognitive Neuroscience.* Psychology Press: East Sussex. 55-63

Dennett, D. 1998. No bridge over the stream of consciousness. *Behavioural and Brain Sciences.* 21: 753-754

Dennett, D., 2002. How Could I Be Wrong? How Wrong Could I Be? *Journal of Consciousness Studies.* 9(5-6): 13-16

Dennett, D. 2007. Heterophenomenology Reconsidered. *Phenomenology and the Cognitive Sciences.* 6: 247-270

Dennett, D. 2015. Why and How Does Consciousness Seem the Way it Seems? In T. Metzinger & J. Windt (eds) *Open MIND.* [Accessed online 2 February 2019] http://dx.doi.org/10.1080/13869795.2013.852615

Dretske, F. 2008. Simple Seeing. In W. Lycan & J. Prinz (eds.) *Mind and Cognition.* Oxford: Blackwell. 731-740

Dummett, M. 1973. Truth. In his *Truth and Other Enigmas.* Cambridge, MA: Harvard University Press. 1-25

Dummett, M. 1975. Wang's Paradox. *Synthese*. 30 (3-4): 201-32

Duncan, M. 2014. A challenge to Anti-Criterialism. *Erkenntnis.* 79: 283-296

Einstein, A. 1905. On the Electrodynamics of Moving Bodies. *Annalen der Physik.* 17: 891-921

Einstein, A., Podolsky, B., & Rosen, N. 1935. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* 47(10): 777—780

Eddington, A. 1928. *The Nature of the Physical World*. Cambridge: Cambridge University Press

Evans, G. 1982. *The Varieties of Reference*. Oxford: Oxford University Press

Fara, D. 2001. Phenomenal Continua and the Sorites. *Mind.* 110: 905-35

Feldman, F. 2000. The Termination Thesis. *Midwest Studies in Philosophy.* 24: 98 – 115

Fine, K. 1975. Vagueness, Truth and Logic. *Synthese.* 30(3-4): 265-300

Gale, R.M. 1968. *The Language of Time.* London: Routledge

Garrett, B. 1991. Personal Identity and Reductionism. *Philosophy and Phenomenological Research.* 51(2): 361-373

Goff, P. 2017. *Consciousness and Fundamental Reality*. Oxford: Oxford University Press

Goodenough, J. 1996. Parfit and the Sorites Paradox. *Philosophical Studies.* 2(2): 113-20

Grice, H. P. 1961. The causal theory of perception. *Aristotelian Society Supplementary Volume.* 35: 121-53

Hensen, B., Bernien, A., Dréau, A. *et al.* 2015. Loophole-free Bell inequality violation using electron spin separated by 1.3 kilometres. *Nature*. 526**:** 682–686

Hoerl, C. 2014. Do we (seem) to perceive passage? *Philosophical Explorations.* 17(2): 188-202

Hoerl, C. 2017. Temporal experience and the philosophy of perception. In I. Phillips (ed.) *The Routledge Handbook of Philosophy of Temporal Experience.* London: Routledge. 171-183

Huemer, M. & Kovitz, B. 2003. Causation as simultaneous and continuous. *Philosophical quarterly.* 53: 556-65

Hume, D. 1740 [1992]. *A Treatise of Human Nature*. New York: Buffalo

Hummel, P. 2017. Against the Complex Versus Simple Distinction. *Erkenntnis.* 82: 363-378

Husserl, E. 1950. *The Phenomenology of Internal Time-Consciousness* (translated by J.S. Churchill). Bloomington: Indiana University Press

Ishmael, J. 2013. Temporal Experience. In C. Callender (ed.) *The Oxford Handbook of Philosophy of Time*. Oxford: Oxford University Press. 460-482

James, W. 1952. *The Principles of Psychology.* Chicago: Encyclopedia Britannica

Johansson, J. 2007. What is animalism? *Ratio.* 20(2):194-205

Kant, I. 1787 [1996]. *Critique of Pure Reason* (translated by N. Smith). New York: St. Martin's Press

Kolers, P. A. & von Grünau, M. 1976. Shape and Colour in Apparent Motion. *Vision Research.* 16: 329-335

Kon, M. & Miller, K. 2015. Temporal Experience: Models, Methodology and Empirical Evidence. *Topoi.* 34:201-216

Langford, S. 2017. A Defence of Anti-Criterialism. *Canadian journal of philosophy.* 47(5): 613-630

Le Poidevin, R. 2007a. Action at a distance. *Royal institute of philosophy supplement.* 61: 21-36

LePoidevin, R. 2007b. *The Images of Time.* Oxford: Oxford University Press

Lee, G. 2014. Extensionalism, Atomism, and Continuity. In L. Oaklander (ed.) *Debates in the Metaphysics of Time*. London: Bloomsbury. 149-173

Lévi, R. 1927. Théorie de l'action universelle et discontinue. *Journal de Physique et le Radium.* 8 (4): 182–198

Lewis, D. 1986. *On the Plurality of Worlds.* Oxford: Blackwell

Lewis, D. 1973. Causation. *Journal of Philosophy*. 70(17): 556–567

Lewis, D. 2004. Void and Object. In J. Collins, N. Hall & L. Paul (eds.). *Causation and Counterfactuals.* Cambridge: MIT Press. 277-290

Locke, J. 1694 [1975]. *An Essay Concerning Human Understanding.* P. Nidditch (ed.). Oxford: Oxford University Press

Mackie, D. 1999. Personal Identity and Dead People. *Philosophical Studies.* 95: 219 -242

Macpherson, F. 2014. Is the Sense-Data Theory a Representationalist Theory? *Ratio.* 27(4): 369-392

Markosian, N. 1998. Simples. *Australasian Journal of Philosophy*. 76: 213-226

Martin, M.G.F. 1994. Perceptual Content. In S. Guttenplan (ed.) *A Companion to the Philosophy of Mind.* Oxford: Blackwell. 463-471

Mazijk, C. 2015. Do We Have To Choose Between Conceptualism and Non-Conceptualism? *International Journal of Philosophical Studies.* 23(5): 645-665

McDaniel, B. 2009. Presentism and Absence Causation: An Exercise in Mimicry. *Australasian Journal of Philosophy*, 88(2): 323–332

McDaniel, K. 2007. Extended Simples. *Philosophical Studies.* 133: 131-141

McDowell, J. 1994. *Mind and World*, Cambridge MA: Harvard University Press

McDowell, J. 1997. Reductionism and the First Person. In J. Dancy (ed.) *Reading Parfit.* Oxford: Blackwell. 230-250

McDowell, J. 2008. Non-conceptual Content. In W. Lycan & J. Prinz (eds.) *Mind and Cognition.* Oxford: Blackwell. 748-59

McGinn, C. 1983. *The Subjective View.* Oxford: OUP

McTaggart, J. 1908. The Unreality of Time. *Mind.* 17(68): 457-474

Mellor, D.H. 1995. *The Facts of Causation.* London: Routledge

Mellor, D. H. 1988. Crane's Waterfall Illusion. *Analysis.* 48(3): 147-150

Merricks, T. 1998. There Are No Criteria of Identity Over Time. *Nous*. 32: 106-124

Merricks, T. 2006. Goodbye Growing Block. In D. Zimmerman (ed.) *Oxford Studies in Metaphysics.* Oxford: Clarendon Press. 103-10

Moore, G.E. 1903. The Refutation of Idealism. *Mind.* 12(48): 433-453

Noë, A. 2007. The Critique of Pure Phenomenology. *Phenomenology and the Cognitive Sciences.* 6: 231-245

O'Regan, K. 1992. Solving the 'real' mysteries of perception: The world as an outside memory. *Canadian Journal of Psychology*. 46: 461-88

Olson, E. 1997. *The Human Animal.* Oxford: Oxford University Press

Olson, E. 2003. An Argument for Animalism. In R. Martin & J. Baresi (eds.) *Personal Identity.* Oxford: Blackwell. 318 - 334

Olson, E. 2012. In Search of the Simple View. In G. Gasser & M. Stefan (eds.) *Personal identity: complex or simple?* Cambridge: Cambridge University Press. 44-62

Parfit, D. 1987. *Reasons and Persons.* Oxford: Oxford University Press

Parfit, D. 1993. The Indeterminacy of Identity: A Reply to Brueckener. *Philosophical Studies.* 70(1): 23-33

Parfit, D. 1995. The Unimportance of Identity. In H. Harris (ed.) *Identity.* Oxford: Oxford University Press. 13-45

Parfit, D. 1999. Experiences, Subjects, and Conceptual Schemes. *Philosophical Topics.* 1/2: 217-270

Parfit, D. 2008. Persons, Bodies, and Human Beings. In T. Sider, J. Hawthorne & D. Zimerman. *Contemporary Debates in Metaphysics.* Oxford: Blackwell. 177 – 208

Pariyadath, V. & Eagleman, D. 2007. The effect of predictability on subjective duration. *PloS One*. 2(11): e1264. [Accessed online 2 July 2018] https://doi.org/10.1371/journal.pone.0001264

Paul, L.A. 2010. Temporal Experience. *Journal of Philosophy.* 107(7): 333-359

Peacock, C. 1992. *A Study of Concepts*. London: MIT press

Pessoa, L., Thompson, E., Noë, A. 1998. Finding out about filling-in. *Behavioural and Brain Sciences.* 21: 723-747

Phillips, I. 2011a. Indiscriminability and Experience of Change. *Philosophical Quarterly*. 61: 808-27

Phillips, I. 2011b. Perception and Iconic Memory: What Sperling Doesn't Show. *Mind and Language*. 26: 381-411

Pollock, H. 2018. Parfit's Fission Dilemma. *Theoria*. 84(4): 284-94

Prior, A. 1959. Thank goodness that's over. *Philosophy*. 34: 12-17

Prosser, S. 2016. *Experiencing Time.* Oxford: Oxford University Press

Prosser, S. 2017. Rethinking the Specious Present. In I. Phillips (ed.) *The Routledge Handbook of Temporal Experience.* 146-156

Reid, T. 1855. *Essays on the Intellectual Powers of Man.* Derby: Boston

Rensink, R. 2000. The Dynamic Representation of Scenes. *Visual Cognition*. 7(1-3): 17–42

Roelofs, L. 2016. The Unity of Consciousness, Within Subjects and Between Subjects. *Philosophical Studies*. 173(12): 3199-3221

Rose, D. & Summers, J. 1995. Duration illusions in a train of visual stimuli. *Perception*. 24(10): 1177-1187

Rothstein, S. 2004. *Structuring Events.* Oxford: Blackwell

Rovelli, C. 2016. *Reality is not what it seems* (trans. S. Carnell & E. Segre). London: Allen Lane

Russell, B. 1903. *The Principles of Mathematics.* Cambridge: Cambridge University Press

Russell, B. 1921. *The Analysis of Mind.* Duke University Press

Russell, B. 1923. Vagueness. *Australasian Journal of Philosophy and Psychology.* 1: 84-92

Sainsbury, M. 1991. Is there Higher-Order Vagueness? *Philosophical Quarterly.* 41(163): 167-182

Schwitzgebel, E. 2007. No unchallengeable epistemic authority, of any sort, regarding our own conscious experience – contra Dennett. *Phenomenology and the Cognitive Sciences*. 60(1): 107-113

Schwitzgebel, E. 2008. The Unreliability of Naïve Introspection. *Philosophical Review.* 117(2): 245-273

Shalm, L., Meyer-Scott, E., Christensen, B., *et al.* 2015. A strong loophole-free test of local realism. *Phys. Rev. Lett*. 115: 250402. [Accessed online 23 October 2019] https://doi.org/10.1103/PhysRevLett.115.250402

Shoemaker, D. 2002. The Irrelevance/ Incoherence of Non-Reductionism About Personal Identity. *Philo*. 5(2): 143-160

Shoemaker, S. 1997. Parfit on Identity'. In J. Dancy (ed.) *Reading Parfit*. Oxford: Blackwell. 135-148

Shoemaker, S. 2012. Against Simplicity. In Gasser, G. & Stefan, M. (eds.) *Personal Identity: Complex or Simple?* Cambridge: Cambridge University Press. 123-136

Sider, T. 1997. Presentism and Ontological Commitment. *The Journal of Philosophy.* 96 (7): 325 – 347

Sider, T. 2007. Parthood. *Philosophical Review*. 116: 51-91

Sider, T. 2011. *Writing the Book of the World.* Oxford: Oxford University Press

Simons, D. & Levin, D. 1997. Change Blindness. *Trends in Cognitive Sciences.* 1(1): 241-82

Smith, Q. 1993. *Language and Time. Oxford: Oxford University Press*

Sorensen, R. 1988. *Blindspots.* Oxford: Oxford University Press

Sorensen, R. 2001. *Vagueness and Contradiction.* Oxford: Oxford University Press

Soteriou, M. 2007. Content and the Stream of Consciousness. *Philosophical Perspectives*. 21: 543-568

Soteriou, M. 2013. *The Mind's Construction.* Oxford: Oxford University Press

Steward, H. 1997. *The Ontology of Mind: Events, Processes and States.* Oxford: Oxford University Press

Swinburne, S. 1984. Personal Identity: The Dualist Theory. In S. Shoemaker & R. Swinburne (eds.) *Personal Identity.* Oxford: Basil Blackwell. 1-66.

Thorson, J., Lange, G. D., and Biederman-Thorson, M. 1969. Objective Measure of the Dynamics of a Visual Movement Illusion. *Science.* 164:1087-8

Tye, M. 2002. Representationalism and the Transparency of Experience, Noûs, 36**:**137-51

Unger, P. 1990. *Identity, Consciousness and Value.* Oxford: Oxford University Press

Van Gulick, R. 1998. Analytical isomorphism and Marilyn Monroe. *Behavioural and Brain Sciences.* 21: 776-777

Velmans, M. 2007. Heterophenomenology versus critical phenomenology. *Phenomenology and the Cognitive Sciences.* 6: 221-230

Williams, J.R.G. 2008. Ontic Vagueness and Metaphysical Indeterminacy. *Philosophy Compass.* 3/4: 763-788

Williams, J.R.G. 2016. 'Vagueness as Indecision'. *Aristotelian Society Supplementary Volume.* 90(1): 285-309

Williamson, T. 1992. Vagueness and Ignorance. *Aristotelian Society Supplementary Volume*. 66: 145-177

Williamson, T. 1999a. On the Structure of Higher-Order Vagueness. *Mind.* 108(429): 127-143

Williamson, T. 1999. Existence and Contingency. *Aristotelian Society Supplementary Volume.* 73: 181-203

Wright, C. 1992. Is Higher Order Vagueness Coherent? *Analysis.* 52(3): 129-139

Zeki, S. 1991. Cerebral akinetopsia (visual motion blindness). A review. *Brain*. 114 (2): 811–24

Zimmeman, D. 1997. Chrisholm and the essences of event. In Hahn, L. (ed.) *The philosophy of Roderick M. Chrisholm.* Illinois: Open Court. 73-100

Zimmerman, D. 1998. Criteria of Identity and the 'Identity Mystics'. *Erkenntnis.* 48(2-3): 281-301

Zimmerman. D. 2011. Presentism and the space-time manifold. In C. Callender (ed.) *The Oxford Handbook of Philosophy of Time.* 163-246