Being Consciousness

A phenomeno-analytical investigation into the relationship between consciousness and selfhood

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by Tom Anders Winfield.

August 2015

Abstract

The notion that we are essentially conscious beings has a good deal of intuitive appeal, but also gives rise to a number of philosophical problems. As a result of its appeal, and in conjunction with a growing dissatisfaction with reductive accounts of consciousness, a number of experiential accounts of personal identity have been introduced into the relatively recent literature. These accounts offer various analyses of the relationship between consciousness and selfhood in an attempt to overcome the problems faced by adopting such a position. I argue that a correct appreciation of the nature of inner awareness, and experience more generally, entails that the experiential approach is indeed justifiable. Specifically, I argue that the relationship between an experience and its subject necessitates the view that selves are constituted by episodes of consciousness. I then evaluate a number of theories of temporal consciousness and argue that the most promising kind of account has implications concerning our persistence conditions. Subsequently, I argue for a radical account of our nature by defending the resulting ontological claim: selves are streams of consciousness.

Contents

Acknowledgements

With thanks to Barry Dainton, Christopher Bartley and Tom McClelland.

This thesis is dedicated to Rosie, Catherine, and my parents.

Introduction

> It is certain there is no question in philosophy more abstruse than that concerning identity, and the nature of the uniting principle, which constitutes a person. (Hume, 2007, p.127)

> I am Seeing, pure and by nature changeless … I am unborn, abiding in Myself. (Śankarācārya, 1992, p.123)

> The conviction which every man has of his identity, as far back as his memory reaches, needs no aid of philosophy to strengthen it; and no philosophy can weaken it, without first producing some degree of insanity. (Reid, 1854, p.241)

What am I? When beings such as ourselves entertain this question a curious friction can arise: on the one hand we have a sense of knowing what the question is about. It is about me - just *this*, myself. It is about this being that I am, that I have been since I can remember, and that I have no practical problem in identifying each and every day: it is so familiar that I don't need to identify it. On the other hand, we can also be met with a deep sense of uncertainty when we try to spell out, precisely, what it is that this "me" *is*. This becomes even more pronounced when we engage in the philosophical task of identifying our *essential* nature and of specifying our persistence conditions - what it takes for this "me" to continue on through time as the self-same being. It seems that the most familiar being there is, one's very own self, gives rise to some of the most difficult philosophical problems. It is both a unique question and a deeply personal question. Its uniqueness is a matter of its direction: as conscious and curious beings we look outwards at the world we find ourselves in and ask questions of it. We ponder and investigate its nature, its cause and its meaning. The problem of personal identity turns such an inquiry back in on itself: who, or what, are *we* who ask these questions? It is a powerfully intimate problem: we can attempt as best we can to retain the role of unbiased investigators, treating the subject matter as one more sphere of inquiry, but any conclusions we reach will quite obviously have the ability to affect us

personally. For all of these reasons it is a philosophical problem worthy of study: it is unique, personal, and challenging. It is also a highly contested area of inquiry: despite the importance of the question, and the powerfully personal context that has motivated an enormous amount of work on the issue for thousands of years, we have not reached a consensus. Unless we take this as evidence of the inexplicable nature of the problem (or, perhaps, even if we do) this means that there is still work to be done. The following inquiry represents one more attempt to solve this puzzle or, more realistically, present the foundations of a new theory. Less ambitiously, it will shed light on a number of the most important issues at play by re-evaluating the implications of some of the strongest options available to us.

It is important to provide clarity on a number of points before undertaking this investigation: the form of the general question under consideration needs to be stated and explained, the terms being utilised need to be clarified, and any relevant background assumptions and commitments need identifying. To a certain extent the first and second issues are interdependent: clarifying one of them elucidates the other. Eric Olson has provided a useful interpretation of the issue:

> What are we? That is, what are we metaphysically speaking? What is our basic metaphysical nature? What are our most general and fundamental properties? I claim that this is a real and important question. It is different from the traditional mind-body problem and from familiar questions of personal identity. It is frequently neglected. And it is fiendishly difficult to answer. (Olson, 2007, p.37)

To clarify:

> The most familiar problems of personal identity are the *personhood question* and the *persistence question*. The personhood question asks what it is to be a person. What is necessary and sufficient for something to count as a person, as opposed to a non-person? What have people got that non-people haven't got? The persistence question asks what it takes for us (or for people in general) to persist through time. What sorts of adventures is it possible, in the broadest sense of the word 'possible', for you to survive? What sort of thing would necessarily bring your existence to an end? What determines which past or

> future being is you? The question of what we are is more or less completely unrelated to the personhood question. What qualifications a thing needs in order to count as a person is one thing; what sort of things actually have those qualifications - organisms, bundles of perceptions, or what have you - is another. (Olson, 2007, p.42, original italics)

His final point above has in fact become the focus of debate in an area outside of specialised philosophical discourse: the growing movement to recognise the personhood of certain nonhuman animals, and thereby afford them their relevant rights, speaks to this issue. We might also see similar developments in the years to come if artificial intelligence research delivers on some of its grander promises: we could see the creation of a self-conscious intelligence that, by any reasonable definition, matches our criteria for personhood. But surely, Olson's argument goes, a nonhuman person, an artificial person, and persons such as ourselves are different *kinds* of beings in some significant sense? Olson claims that to know "what it is to be a person is therefore not to know what we are. Likewise, to know what we are is not to know what it is to be a person" (Olson, 2007, p.42). Of course, it follows straightforwardly that if we are *essentially* persons then to know what it is to be a person is to know what we are. Olson's point, however, is that if there is more to be said about what we are - metaphysically considered - then identifying the necessary and sufficient conditions for personhood will not be enough.

A similar point seems to be reasonable with regards to the typical form of the persistence question:

> What it takes for us to persist may be one aspect of our metaphysical nature. Knowing our persistence conditions would tell us *something* about what sort of things we are. But it wouldn't tell us much. An account of our persistence conditions would not by itself tell us whether we are material or immaterial, or what parts we have, or whether we are substances. What it takes for a person to persist through time is one thing; what sort of beings *have* those persistence conditions, or indeed whether any do, is something else. (Olson, 2007, p.43, original italics)

Olson could be accused of being a little too pessimistic with regards to the elucidation

that might be achieved in settling such a matter. It is also not obvious that the basic question concerning what we are is "*more or less completely* unrelated to the personhood question" (Olson, 2007, p.42, my italics). But it seems right to agree with him that the question of what kind of beings we are, metaphysically considered, is different to the usual questions of personal identity when they are framed in the ways he identifies. For Olson,

> That may explain why it is often neglected. It is common practice to defend an account of our identity over time at great length without saying a word about what we are, except perhaps to rule out our being immaterial substances … When the matter is addressed at all, it is frequently little more than an afterthought … (Olson, 2007, p.44)

Taking our lead from Olson we will begin our inquiry by confronting this more general question first. Having done so, we will then turn our attention to the issue of personal persistence. To the general question of what kind of beings we are the following (partial) answer is obviously true: we are conscious beings. Whatever else may need to be said it is a fact that sometimes we are conscious: indeed, it is only by being conscious and knowing this that we can engage in a self-directed inquiry of the kind we are considering. It seems, then, an inevitable starting point from which to begin answering the general question of our nature. In focusing down on the question of what it is to be a conscious being, and, specifically, in dealing with the relationship between a subject and its experiences, we will find implications that inform the persistence question (both in terms of the form of the question and its most promising answer). We will see that Olson's suggestion to pay attention to the general question of our nature is highly useful: there are good reasons to think that only a certain *kind* of being can be conscious in the way that we are, and that such a consideration proves damaging to a host of personal identity theories.

Given, then, that we will take the relationship between a subject and its experience as our starting point, it is important to be clear on the general picture of experience assumed here. If there were just one philosophical problem more "abstruse", as Hume puts it, than that of personal identity then the hard problem of consciousness would be a solid candidate. The relationship between mind and matter is a notoriously difficult question and is itself another area in which a consensus has yet to be reached. As

such, it will not be the target of this investigation. Many of the following claims can in fact be appropriated into a number of different theories of consciousness and to a large extent this work will remain suitably agnostic on the issue. There are, however, some general assumptions that will be made, the most important of which is that conscious experience is *real* and therefore something we will take seriously. In the coming sections the work of Barry Dainton will be covered significantly: his theory of personal identity deals with the relationship between a self and its experiences in an uncommonly substantial way and therefore has much in common with this forthcoming account. As will become clear there are important differences between Dainton's *Phenomenal Self* and the *Ephemeral Self* to be presented here, but the following stance is shared:

> By 'taking experience seriously' I mean adopting a stance of robust, full-blooded realism about consciousness. This means taking consciousness as seriously as we take science. From this perspective, sensory experiences, bodily sensations and conscious thoughts are regarded as just as real as paradigmatic physical things such as mountains, houses and trees, and perhaps more real than some of the currently postulated occupants of the microphysical realm. It also means rejecting all attempts to reduce the experiential to the non-experiential. (Dainton, 2006, p.1)

Another realist about experience, Galen Strawson, puts it the following way:

> I say that I'm a *real* realist about experience because some who claim to be realists about experience aren't really any such thing. What do I mean by real realism about experience? The quickest way to say what it is is to say that it's to hold exactly the same general view about what experience is (colour experience, say, or pain experience, or taste experience), considered specifically as experience, that one held *before one did any philosophy*, e.g. when one was thirteen or ten or six. One then had an entirely correct view. If people ask what that view is I'll ask them to think back to their childhood. If they say they still don't know I won't believe them. (Strawson, 2015, forthcoming, original italics)

With regards to the mind-body problem and the shift towards a physicalist account of consciousness that has arisen in recent decades, the claims of the following sections will be committed only to Strawson's fundamental point here:

> You're certainly not a realistic physicalist, you're not a real physicalist, if you deny the existence of the phenomenon whose existence is more certain than the existence of anything else: experience, 'consciousness', conscious experience, 'phenomenology', experiential 'what-it's-likeness', feeling, sensation, explicit conscious thought as we have it and know it at almost every waking moment. Many words are used to denote this necessarily occurrent (essentially non-dispositional) phenomenon, and in this paper I will use the terms 'experience', 'experiential phenomena' and 'experientiality' to refer to it. (Strawson, 2006, p.3)

Although I will remain agnostic on the wider issue, some of the claims to be presented here might have implications for the debates concerning the relationship between mind and matter. Panpsychism is a position that is now being taken more seriously than in previous years. The issue of the relationship between experiences and subjects plays an important role in the current literature in this area and the thesis presented here speaks directly to this relationship.[1]

I will follow Strawson in treating "consciousness" and "experience" as, ultimately, synonymous. I will also employ the popular terminology used by Dainton:

> By 'consciousness' I mean *phenomenal* consciousness; by 'experiences' I mean states or items with a phenomenal character. The 'phenomenal character' of an experience refers to the distinctive feel the experience has. A state has a phenomenal character when there is something that it is like to have or undergo that state. (Dainton, 2006, p.2, original italics)

Consider your current experience. You are conscious of, among other things, the words you are currently reading. This experience has a certain "feel" to it: there is something it is like for you to be doing what you are currently doing. What it is like

[1] See Coleman (2013) and Goff (2006) on this particular issue.

for you to be reading this is different from what it would be like for you to be swimming in cold water, or watching a film, or doing exactly what you are now doing but with a headache (assuming you don't already have one). The distinctive feel of your current experience just *is* its phenomenal character. Crucially, this experience is the one thing whose existence you can be most certain of during its occurrence. It seems possible that the world in which we live is a computer generated virtual reality. This, of course, strikes us as unlikely but it is perfectly *conceivable*. Your experience, however, considered just in its experiential "what-it's-likeness" cannot be an illusion in this sense: the appearance-reality distinction is not applicable to phenomenal consciousness due to the fact that phenomenal consciousness *is* the appearance of this seemingly external world (among other things). There is a certain sense in which if someone is to claim that they do not recognise the target of the term "experience" in the way just described there is little that can be done. Perhaps the most I can do is exclaim, "It is *this*" and gesture openly at nothing in particular. As well as assuming the reality of experience, I will assume that the target of the above pointers is recognisable for anyone who genuinely engages with them.

I will also follow Strawson in his treatment of the concepts of "experience", "subject", and "subjectivity" in the following sense:

> Experience is necessarily experience-*for* - experience for someone or something. I intend this only in the sense in which it's necessarily true … To claim that experience is necessarily experience-for, necessarily experience-for-someone-or-something, is to claim that it's necessarily experience on the part of a subject of experience … Some say one can't infer the existence of a subject of experience from the existence of experience, only the existence of subjectivity, but I understand the notion of the subject in a maximally ontologically non-committal way - in such a way that the presence of subjectivity is already sufficient for the presence of the subject, so that 'there is subjectivity, but there isn't a subject' can't possibly be true. (Strawson, 2011, pp.274-5, original italics)

And, finally, I will follow Dainton's policy with regards to individuating experiences:

> I will assume that token experiences owe their individuality to three factors:

> their exact phenomenal character, their time of occurrence, and their physical basis. In keeping with my stance of moderate naturalism, I will not speculate exactly what form this physical basis takes. (Dainton, 2006, p.25)

The scientific picture of our place in the world will also be taken seriously: as organisms we have evolved through a process of natural selection and, although an agnostic stance with regards to the ultimate relationship between mind and matter is adopted here, it is reasonable to assume that this process of evolution has also been responsible for the development of our mental capacities. Towards the end of our investigation we will consider some soteriological aspects of the theory on offer. We will find agreement with particular schools of Indian Philosophy in their claim that recognising certain facts pertaining to our own nature, and attending to specific features of our experiential makeup, can facilitate meaningful and beneficial psychological change. I think this can be argued for without straying from a broadly naturalistic perspective concerning our development as a species (and without denying the existence of the self). Speaking generally for now, the seemingly universal impulse for salvation can be met without recourse to anything supernatural, eternal or in any other way out of step with our current scientifically informed picture of reality.

The structure of the thesis is as follows:

§1: *Inner Awareness and The Novelty Problem*

The following conclusion will be argued for: whenever we are conscious, we are aware of our experience, i.e. inner awareness is ubiquitous. This claim will be justified by considering a host of serious problems that face rival views. The most popular alternative theories fall into two kinds: the first claims that we are never aware of the intrinsic nature of our experiences and the second claims that we are not *usually* aware of our experiences, but can become so through introspection. The Novelty Problem will be presented and shown to be damaging to both of these views.

§2: *Higher-order Theories and Inner Awareness*

The nature of this inner awareness will be investigated. I will argue that inner

awareness is not the result of a so-called “higher-order” awareness of any kind, and that experience is self-intimating.

§3: *Being Consciousness*

The Exclusivity Argument will be presented. Employing the conclusions of §1 and §2 as premises it yields the (preliminary) *Identity Thesis*, which holds that the subject of an experience (*e*) is identical with *e*.

§4: *The Stream of Consciousness*

Having established a preliminary answer to the general question of our nature, I will then outline the most promising account of temporal consciousness and thereby present a theory of our persistence conditions. The findings here will motivate a reinterpretation of *The Identity Thesis*, resulting in *The Exhaustive Constitution Thesis*.

§5: *The Ephemeral Self*

The conclusions for §3 and §4 will result in *The Stream Thesis* which holds that the subject of experience *x* is identical with the stream of diachronically co-conscious experiences of which *x* is a part. The use of thought experiments within the literature of personal identity will be partially justified and then employed to strengthen both the believability and soundness of *The Stream Thesis*.

§6: *The Personal Copernican Revolution*

The reconceptualisation of death entailed by *The Stream Thesis* will be investigated and its related soteriological possibilities will be evaluated with reference to similar claims made in certain schools of Indian Philosophy.

1: Inner Awareness and The Novelty Problem

§1.1 The landscape

Consider your current experience. Through it you are conscious of these words and their (more or less) central location in your visual field. Although you are focused on this text you are also conscious of the page on which it appears and, moving further out from your focal point, you are conscious of less clearly-defined yet still visually present objects: keep your head still but move your eyes to focus on whatever is just above the top of this page in your visual field - a table, the floor, a wall - and then refocus on this sentence: the background object is certainly less vivid now, but it is still visible; the area above the top of this page is not a visual void. Consider, too, the sounds you are currently hearing and the feeling of your body in your seat. Through your current experience you are conscious of all these things and more. In short, your experience makes you conscious of the world. At this point we can ask an important question: are you also conscious of the experience itself? If the answer to this is affirmative then a further question must be addressed: were you conscious of your experience before you were encouraged to attend to it?

It is a fascinating fact that these simple questions about that most familiar of things (our very own lived experience) have elicited such markedly opposing answers within the literature. The metaphysics of consciousness, particularly the attempt to find a place for conscious experience within our current scientifically informed picture of reality, has led to infamously disparate positions: for every bullet-biting eliminativist who proclaims that although there seems to be phenomenology "it does not follow that *there really is* phenomenology" (Dennett, 1993, p.366, original italics) we can find an equally ardent defender of experience as "the fundamental given natural fact" (Strawson, 2006, p.4). But this mix of radically opposed views is perhaps somewhat understandable given the sheer intellectual difficulty we face in reconciling what have come to be known as the scientific and manifest images (Sellars, 1963) (an effort that may indeed require a radical overhaul of our current thinking on the issue (McGinn, 2004)). One might hope that, in contrast to such fundamental disagreement, we might be able to reach something approaching a consensus on at least *some* aspects of

experience if we could just put the mind-body problem on the back burner and focus on experience *itself* as opposed to where and how it fits into the rest of the world. After all, conscious experience *as lived* is the most familiar thing there is.

But things aren't quite so easy: even these two simple questions about the everyday nature of experience have given rise to fundamental disagreement. The reasons for this are likely manifold and interconnected; chief suspects are the limits of discursive thought (Strawson, 2013, p.28), prior ontological commitments (not everyone can, or will, leave the mind-body problem on the back burner) and confusion resulting from an absence of standardised concepts in the literature. This is particularly true for the term "self-consciousness". It seems that despite its eminent familiarity the nature of our very own lived experience is not as obvious as we might have thought it. It is tempting to wonder whether we are in some sense *too* close to it: it is as if we cannot hold it at arm's length to get a good look at it. But such pessimism should be resisted and, as we will see, so long as care is taken to avoid certain stubbornly recurring conceptual traps we can indeed arrive at some reasonably confident conclusions concerning the nature of conscious experience as it is in itself.

It is helpful to approach this issue of inner awareness, as it has come to be known, by breaking it down into two questions: firstly, are we ever directly conscious of experience and if so, secondly, is this only when we are explicitly attending to it? A rough picture of the landscape of answers has three main camps: firstly, those who argue that inner awareness is a ubiquitous feature of experience (i.e. we are always conscious of our experience when we are conscious at all, regardless of whether or not we are explicitly attending to it), secondly - in stark opposition - those who argue that we are never directly conscious of the intrinsic qualities of experience even when we are engaged in what would colloquially be called "attending to experience" (i.e. experience is, in some sense, transparent) and, thirdly, those who inhabit a middle-ground between these views and hold that although we are not usually conscious of our experience we can become so through introspection. Each camp has several different ways of fleshing out their shared line but a general picture of the current debate sees these as the available options.

As it turns out, there are good reasons to think that there are in fact only two internally consistent options open to us. A middle-way view can admittedly be an attractive one: it promises to account for our ability to introspect our experiences whilst at the same time avoiding the (for some) counter-intuitive idea that experience

and inner awareness are inseparable. In short, it accepts what many take to be the most basic of phenomenological facts (that we can, at least sometimes, be directly conscious of the intrinsic character of our own experiences) without positing a *ubiquitous* inner awareness (which hereafter will be referred to as "UIA"). There are reasons to think that this turns out to be an empty promise: some of the strongest middle-way views gain their plausibility in large part through their willingness to pay due respect to our phenomenology but, as we will see, find themselves committed to positing either an implausible degree of transparency (with regards to experience) or a ubiquitous inner awareness in order to remain consistent. Although a *fully* comprehensive survey of the different options available to a middle-way theorist will not be presented here, we will see that a number of the most plausible versions of the middle-way view collapse into transparency or UIA views. If we take this to be indicative of the wider situation then we are left with just two general options when considering the questions posed above: either hold that inner awareness is a ubiquitous feature of experience or defend some kind of transparency view. Transparency views, however, face their own problems. The notion of transparency, then, is crucial to the debate and care must be taken to understand it as clearly as possible.

§1.2 Transparency and the perceptual model of introspection

On this issue of transparency:

> It … has appeared obvious to some philosophers that the so-called transparency of experience supports the following claim: either our experiences do not have an intrinsic phenomenal character or we are unable to attend to these intrinsic features. (Nida-Rümelin, 2007, p.429)

As we shall see, however, not everyone accepts such an entailment. The first use of the metaphor in relation to experience is usually attributed to G. E. Moore (1903), but its role and meaning alter depending on the theory it is embedded in and, subsequently, the transparency claims relevant to the present discussion differ from Moore in their application of it. Nevertheless, it is helpful to be clear about its origins:

if we find the current use of the metaphor to be unsupportable we may still find some use for it, perhaps by returning to its original meaning (or something like it). This is Moore's introduction of the term:

> The term 'blue' is easy enough to distinguish, but the other element which I have called 'consciousness' - that which sensation of blue has in common with sensation of green - is extremely difficult to fix. That many people fail to distinguish it at all is sufficiently shown by the fact that there are materialists. And, in general, that which makes the sensation of blue a mental fact seems to escape us; it seems, if I may use a metaphor, to be transparent - we look through it and see nothing but the blue; we may be convinced that there is *something*, but *what* it is no philosopher, I think, has yet clearly recognised. (Moore, 1903, p.446, original italics)

It is clear enough already that Moore, unlike many contemporary defenders of transparency, is not arguing that we are unable to become conscious of the intrinsic character of experience itself.[2] He is merely stating the difficulty faced by someone who tries to describe the common factor shared by all experiences. It is admittedly a simple linguistic step from Moore's view to a more radical transparency (the view that we are *never* directly aware of the intrinsic qualities of experience) - for Moore when we experience blue we "see nothing but the blue" and for the "radical transparency" defender when we experience blue we are *aware* of nothing but the blue (where both cases presuppose a minimally simple experience of blue). But it is not an obviously *correct* step and it is not one made by Moore. He goes on:

> the moment we try to fix our attention upon consciousness and to see *what*, distinctly, it is, it seems to vanish: it seems as if we had before us a mere emptiness. When we try to introspect the sensation of blue, all we can see is the blue: the other element is as if it were diaphanous. Yet it *can* be distinguished if we look attentively enough, and if we know that there is something to look for. My main object in this paragraph has been to try to make the reader *see* it: but I fear I shall have succeeded very ill. (Moore, 1903, p.450, original italics)

[2] See Kind (2003) for a detailed examination of the differences between Moore's view and this more radical conception of transparency.

In fact he goes further and holds that to be aware of a sensation of blue is "to be aware of an awareness of blue; awareness being used, in both cases, in exactly the same sense" (Moore, 1903, p.449). For Moore, then, his characterisation of consciousness as transparent is suggestive rather than explicitly descriptive; in the above selections he tentatively claims that it "*seems* … to be transparent" and that it is "*as if* it were diaphanous" (my italics). The concept has morphed into something much stronger in certain areas of the current debate: experience doesn't just *seem* transparent - it really is. On this view it is not just difficult to be directly aware of the intrinsic qualities of experience; it is impossible. Before moving on to assess the strength of this general view it is worth mentioning Siewert's warning that "[it] is not clear there is some single, unambiguous, literal thesis that encapsulates what authors generally have in mind when they speak of the transparency of experience or consciousness" (Siewert, 2004, p.17). That being said, we can discern a general claim that is shared by the main defenders of this more radical conception of transparency as this: we are never directly conscious of the intrinsic nature of our own experiences. The details of if and how we can, at least in some sense, be conscious of our experiences make for the differing theories, but this general negative claim is helpful to have in mind when considering the landscape of options in the debate about inner awareness. Oftentimes, the arguments for transparency concentrate on the nature of perceptual experience but their scope can be intended for much more, as one of the chief proponents of the transparency thesis, Michael Tye, recently made clear:

> Qualia realism is the thesis that experiences have intrinsic features that are non-intentional and of which we can be directly aware via introspection … Qualia realism is inconsistent with transparency. (Tye, 2014, p.41)

It seems appropriate to label this concept "radical transparency" (hereafter "RT"). This is in contrast to more modest transparency claims having to do with the *difficulty* of attending to consciousness as it is in itself, or the simple and non-committal view of experience as something we metaphorically "see through" to the world beyond our senses. Tye defines transparency as follows:

> As far as awareness goes, the thesis is that when we try to introspect a visual experience occurring in normal perception, we are not aware of the experience or its features (intrinsic or not) period. This, I take it, is the basic thesis of transparency. (Tye, 2014, p.40, original underlining)

Such bold claims require equally strong evidence. This is to be found, according to the transparency defenders, in our very own phenomenology. We are invited to introspect our experience and notice that a full and accurate catalogue of the features and qualities we are aware of will have no room for any intrinsic qualities of the experience itself (though we might be aware *that* we are having an experience of some kind). In general we are never aware of the intrinsic qualities of experience and, specifically, when introspecting a perceptual experience we are not aware of the experience itself at all: it is fully transparent. For some, no doubt, such a view seems so obviously wrong as to warrant no further consideration. However, there are good reasons not to be quite so rash. Firstly, declaring a view "obviously wrong" and then ignoring it does not make for much of an argument. Secondly, amidst the highly counter-intuitive claims of the RT thesis there does seem to be something of worth: it seems reasonable to claim, as Tye does, that upon experiencing such qualities as blueness and roundness "you do not experience your experience as blue or round" (Tye, 2002, p.138). It certainly strikes the ear awkwardly to describe an experience as itself being "round". Thirdly, tackling the issue and explaining exactly how and why the transparency defender mistakenly construes our phenomenology might expose conceptual traps relevant to other debates concerning experience. This final point is arguably the most important and, as will become apparent, speaks to a mistaken way of thinking about experience that repeatedly results in incorrect views concerning consciousness (and, by extension, its relation to selfhood).

For Tye, when we attend to how things look to us - "as opposed to how they are independently of how they look" (Tye, 2003, p.139) - we are introspecting. When we introspect a visual experience the only "particulars" we are aware of are external ones: surfaces, for example, and how they appear (Tye, 2003, p.139). On Tye's account "[we] are not aware of those objects *and* a further inner object or episode" (Tye, 2003, p.139) and the only qualities we are aware of are the ones the *surfaces* seem to have:

> Your experience is thus transparent to you. When you try to focus upon it, you 'see' right through it, as it were, to the things apparently outside and their apparent qualities … By being aware of the external qualities, you are aware of what it is like for you … So, your awareness of phenomenal character is not the direct awareness of a quality of your experience. Relatedly, the phenomenal character itself is not a quality of your experience to which you have direct access. (Tye, 2002, p.139, original italics)

Nida-Rümelin has forcefully argued that in defending such a radical version of the transparency metaphor Tye and others are drawing on a further metaphor: the "perceptual model of phenomenological reflection" (Nida-Rümelin, 2007, p.429). She claims that the arguments of the transparency defenders are oftentimes directed against this particular view, and they implicitly assume that such a (problematic) model is the only one available to someone who claims that we can be conscious of the intrinsic character of our own experiences. According to the perceptual model of phenomenological reflection, when someone attends to the phenomenal character of her experience she "concentrates her attention upon the experience that appears to be there within some inner space and she concentrates her attention upon its apparent qualitative properties, upon its quasi-color or 'mental paint'" (Nida-Rümelin, 2007, p.446). Nida-Rümelin readily accepts that neither Tye nor any of the other transparency theorists explicitly describe the phenomenological reflection of their opponents in this manner, but argues that nevertheless it is an apt "description of a metaphor, of a picture that people have in the backgrounds of their minds … [that] does influence people's intuitions" (Nida-Rümelin, 2007, p.446).[3] We can see this at work in Tye's description above when he claims that, as we introspect, we "are not aware of those objects *and* a further inner object or episode" (Tye, 2002, p.139, original italics). Set up in this way Tye's position seems reasonable: we are not greeted, when introspecting our visual experience of a tree, with some "inner" tree-like object *in addition* to what we perceive as the-tree-out-there. Our overall experience is, considered phenomenally, very much as it was prior to introspection (albeit with a focussing of attention, a reconceptualisation or (particularly when

[3] A "picture in the background of one's mind" is, interestingly, a perceptual metaphor for the perceptual-model metaphor under scrutiny. It is not clear whether or not Nida-Rümelin is purposefully drawing our attention to the ease with which we employ this kind of metaphor when thinking about experience, or if she herself has unwittingly used it; either way its seductive power is nicely illustrated.

engaging in phenomenological reflection for philosophical purposes) new thoughts concerning what one is currently doing). In fact, that there is no grand change in our overall experience upon introspection is a claim held in common by UIA defenders and RT theorists alike: it seems to be one of those rare basic phenomenological facts that both sides take as given.

That the transparentists are indeed arguing against the perceptual model of introspection is evidenced in their presupposing a fundamental shift in the *direction* of attention as opposed to a change in focus. Nida-Rümelin (2007) catalogues examples from various transparentists illustrating such a move including this from Tye himself:

> Suppose you are facing a white wall, on which you see a bright red, round patch of paint. Suppose you are attending closely to the color and shape of the patch as well as the background. Now turn your attention from what you see out there in the world before you to your visual experience. Focus upon *your awareness of* the patch as opposed to *the patch* of which you are aware. (Tye, 1997/2003, Section 6, original italics)

Clearly, Tye is presupposing that his opponents are committed to a distinction between our seeing-something-out-there and our visual experience in his (rhetorical) suggestion to turn our attention from one to the other: the suggested change in direction only makes sense if introspection is a perception-like movement from "external" objects to "internal" ones. The implicit argument in Tye's formulation presents us with a choice: either defend the existence of these further inner-objects or accept that we are never aware of the intrinsic qualities of experience itself, bearing in mind the obvious fact that when we introspect there is no substantial phenomenological change in our overall experience. But, as Nida-Rümelin correctly points out, we do not need to make such a choice if we do not hold introspection to be *perception-like*.

For Tye the choice seems to be between consciousness-of-objects or no consciousness at all and therefore, given that we do not experience "inner-objects" in addition to "outer" ones when attending to our experience, we must not be directly

conscious of our experience at all.[4] But neither the inner awareness defenders nor the middle-grounders need make this choice if they do not hold introspection to be perception-like in the way Tye suggests. There is good reason to think that the actual choice that the notion of transparency correctly identifies is in fact the following: given that we notice no substantial phenomenological change in our overall experience upon introspection either we are *never* directly conscious of experience or we are *always* directly conscious of experience (when we are conscious at all). This is an important point that will be best developed and defended by considering some middle-ground views that attempt to deny this choice by holding that we can *sometimes* be aware of the intrinsic qualities of our experience. Upon seeing this third option as unviable we can then return to the crucial dichotomy and consider which side to take.

§1.3 Unconscious phenomenology and The Novelty Problem

So, as I have set up the issue, a middle-ground view defends the following claims: although we are not normally directly conscious of our experience we can become so by reflecting on it (i.e. introspecting). A middle-grounder is not merely saying that we are not normally *focussed* on our experience (this is, after all, perfectly compatible with the idea that inner awareness is a constant feature of experience) but that we are quite literally *unconscious* of our experience, except when actively reflecting on it. The problem for this view, as has been hinted at above, is that it does not seem to each of us, upon introspecting our experience, that anything particularly new has been revealed to us. There is a *lack* of phenomenological novelty. For brevity's sake, from here on out this will be referred to as "The Novelty Problem".[5] What is notable about this claim is that it is shared by strikingly opposed theories: both RT and UIA defenders make use of this phenomenological data in working towards their conclusions (and it is in fact an integral part of both of their arguments).

[4] This impression is rooted in what Michel Henry dubbed "ontological monism": "the assumption that there is only one type of manifestation, only one type of phenomenality" (Zahavi, 1999, p.226).

[5] For a similar problem see Kriegel's "Argument from Surprise" (2009): when we introspect we are not *surprised* by the content we find. I think that framing the argument in terms of the *reason* for this lack of surprise (namely; that there is a lack of novel phenomenal content) helps to present a more powerful case.

This point stands in favour of the claim when we bear in mind the notorious disagreements that can come from using phenomenological data as evidence for a particular view: a stalemate can quickly be arrived at which sees each side simply disagreeing with the proposed phenomenology of the other, with no clear way to settle the dispute; it's not as if we can enter into our opponent's stream of consciousness to verify their description of it. So when such opposing views share a phenomenological claim it is reasonable to take this as a good indication of its reliability. Further to this it is a claim that can be quite easily confirmed personally. (Despite being about our own experience not all phenomenological claims are easily verified in this way.) Given the kinds of questions being asked in this debate (in other words, given that we cannot verify these sorts of claims in the same way that we successfully can with, for example, (many of) the claims of hard science) this level of broad agreement and personal confirmability affords us the strongest degree of confidence that we can reasonably hope to achieve in such a domain.

But why should this lack of novelty be a problem for middle-ground views? Simply put, because it is reasonable to claim that we should expect to find something interestingly novel, phenomenally speaking, when introspecting if we had indeed moved from the state of being-conscious-of-non-experiential-things-only to the state of being-directly-conscious-of-experience. To appreciate the pull of this expectation remember that we are supposed to be moving from an ordinary awareness of our everyday world to an awareness of the phenomena responsible for, among other things, the hard problem of consciousness (Chalmers, 1996), the explanatory gap (Levine, 1983) and the intuition of distinctness (Papineau, 2002)![6] It seems reasonable to expect that, upon such a proposed change in the contents of our experience, we would be met with at least *some* kind of substantial phenomenological novelty.

At the very least, we can say that the onus is on the defender of such a middle-ground view to explain why there would be none. For the UIA defender there is a ready explanation: we were *already* conscious of our experience (albeit perhaps in some kind of non-focal way), hence nothing new is discovered when reflecting. The RT defenders also have their answer: we are *never* directly conscious of the intrinsic qualities of experience so we shouldn't expect to find something new when we attend

[6] The "intuition of distinctness" is Papineau's term for the psychological difficulty we experience when trying to accept the identity of conscious states with material states: they just strike us as such different kinds of things. For Papineau, to explain this mistaken intuition is to solve the hard problem of consciousness (Papineau, 2002, p.3).

to what our experience is like for us. Ultimately, a middle-grounder who claims that we are not usually conscious of our experience, but can sometimes become so, must either deny the lack of novelty or offer an explanation as to why we are not met with any upon introspection.

Understandably, the most notable attempts are of the second kind. But, as we shall see, such attempts run the risk of moving away from the middle-ground and into either a UIA or a RT view. A particularly popular tactic in attempting to explain this lack of novelty has been to appeal to a special kind of *availability* with regards to our experiences. On this view the reason for the lack of novelty upon introspection is not that we were already conscious of our experience but that our experience was, in some relevant way, *available* to us at any given moment should we have decided to introspect at that particular time. Thus, the lack of novel content is explained by the fact that although we were unaware of the experience prior to reflection, it was not *completely* absent from us: we had unconscious knowledge of it. This is how Thomasson describes her position on the issue:

> Nonetheless, the view I am proposing does preserve the grain of truth behind the common association of conscious states with those we are conscious of. A mental state is made conscious by a phenomenology that ordinarily makes us aware of things in the world around us. Although that phenomenology is not ordinarily the focus of our attention, as an immediate part of that conscious state it is already and automatically *available* so that we can turn our attention to it if we so desire, and gain an awareness of the character of our conscious mental states (in the way that we cannot gain direct awareness of unconscious mental states that lack that phenomenology). That is to say, if I consciously see an orange tree, that mental state has a phenomenological character that makes it seem to me that there is an orange tree there. Ordinarily, my focus is on the orange tree, but since there is also a phenomenology, I can turn my attention to that and examine what my experience of seeming to see an orange tree is like. So the grain of truth of this view is that conscious experiences are those that are *available for* direct introspection; we can focus attention on the phenomenology if we so choose. (Thomasson, 2000, p.205, original italics)

In contrast to Thomasson, UIA views all share the claim that when having an experience a subject is necessarily conscious of that experience. The exact nature of this awareness is explained and described in a number of different ways: some claim that it is a perception-like awareness or a higher-order representation of some kind, and others that it is the result of a pre-reflective consciousness intrinsic to the experience itself. The latter view is of particular importance when considering the position described by Thomasson above.

UIA theorists of the pre-reflective sort can (and should) readily accept Thomasson's claim that our "phenomenology is not ordinarily the focus of our attention" and that we "can focus attention on the phenomenology if we so choose". (It is important to bear in mind that a UIA theorist need not argue that we are conscious of our experience in exactly the same way that we are ordinarily conscious of worldly objects: no one need argue that we "see" our visual experience, for example.) As we have seen, for the UIA theorist the very reason that we are easily able to focus on our phenomenology (and are not met with novelty upon doing so) is that we were conscious of it to begin with; it was already phenomenally there for us, and simply required an alteration of attention in order to come into explicit focus. Thomasson's explanation is not so straightforward. For one thing, on this view someone can be unconscious of his or her current conscious state, despite the fact that it is the qualitative character of this very conscious state that allows them to be conscious of what that state represents in the world:

> on this view it's not that the internal phenomenological character is unconscious; on the contrary, it is the immediately present character that makes it like something to be in that mental state, and enables us to be aware of other things. (Thomasson, 2000, p.204)

The phenomenological character of the conscious state is "immediately present" yet something the subject fails to be conscious of. It is true that "immediate presence" can be meant in such a way that does not entail its referent being something we are conscious of: in a sense the back of my skull is immediately present but I am not currently conscious of it (at least, not until I thought about it as an example). At a stretch we might say that my unconscious beliefs are immediately present in the sense that they are able to affect my current behaviour, and can be available to conscious

introspection. But these uses of "immediate presence" do not seem to be viable for phenomenological character if we take seriously the claim that this character is responsible for making the subject conscious of other things. Let's consider Thomasson's example of seeing an orange tree:

> Ordinarily, my focus is on the orange tree, but since there is also a phenomenology, I can turn my attention to that and examine what my experience of seeming to see an orange tree is like. (Thomasson, 2000, p.205)

We have here a subject, an experience and an *apparent* object (as this may not be a case of veridical perception). From the subject's point of view it is impossible to know with absolute certainty whether or not she is hallucinating an orange tree or actually perceiving one, but either way it *seems* to her that there is one standing on the hill in front of her. What is making something seem to her anyway at all is her phenomenology, i.e. the qualitative character of her experience.[7] That the phenomenology is what makes anything seem anyway at all is a point Thomasson agrees with and is right, I think, to do so. The problem with this picture is that it is difficult to see how qualitative character could be the kind of thing that could do the work of making something *seem* a certain way to a subject whilst itself being *entirely absent* from the conscious apprehension of that very subject. The point can be put like this: a phenomenology of which the subject is completely unconscious is an unconscious phenomenology, and an unconscious phenomenology is *not a phenomenology at all*. Thomasson stresses, however, that the phenomenology is not unconscious - it's just not something the subject is conscious *of*. The trouble is that to be "phenomenal" in this sense standardly means to *appear*, i.e. to be manifest in a subject's consciousness; to be something the subject is conscious of (but, of course, not necessarily to appear in the way an object "out there" does). Thomasson's issue with this way of describing the situation is with the possible implications of "of". There is something suspicious, it seems, with saying that a subject is conscious *of* their phenomenology. She is right to be wary: we are not conscious of our phenomenology in exactly the same way that we are conscious of objects in our

[7] By "qualitative character of experience", recall, I mean the totality of what is phenomenally present for the subject of that experience. This, in turn, determines what it is like for the subject to be that subject at that time.

environment. However, it seems entirely reasonable to hold that the phenomenal content of an experience is something the subject is conscious *of*, no matter how awkwardly we might hear that expression at first glance.

Taking care to note that the following metaphor is not entirely analogous, consider a painting of Thomasson's orange tree. Suppose that it is a very good painting and that it can, in some sense, make it seem to you that there is an orange tree in front of you.[8] If this was a particularly powerful illusion your attention might become explicitly focused on the tree seemingly "over there": you may be able to forget that what is in front of you is in fact a painting. In this sense you are not currently conscious of the painting *as* a painting. But you are still just as conscious of it as you previously were: it is still part of the phenomenal content of your consciousness. If you weren't at all conscious of it, there would not seem to be an orange tree before you. This is because your awareness of the apparent orange tree is (in part) *constituted* by your awareness of the painting.

The same general point can be applied to the case of our phenomenology when perceiving a real tree. Whatever might need to be said about external content, it is right to say that a conscious perception of a tree is in large part constituted by the experiential-qualitative character of that very perception, assuming that direct realism is false. This is precisely *why* it is our phenomenology that "enables us to be aware of other things" (Thomasson, 2000, p.204). No conscious awareness of qualitative character - no experience. No experience - no *conscious* perception. The phenomenological-representational-event (the experience) has to be something we are conscious of if we are to apprehend the tree in the first place. It is true that we may not, prior to introspection, be considering our phenomenology *as* phenomenology (we are not necessarily employing any concept of "experience", just as we may not be employing any concept of "painting" when transfixed by what a picture represents) but we are still, in some sense, conscious of it. "In some sense" needs spelling out and the precise nature of this inner awareness is a difficult matter (there are a number of candidate answers) - but the claim that qualitative character needs to be something we are conscious of in order for it to make us conscious of some object in the world

[8] This is not generally what occurs when we appreciate a painting, of course, but it seems possible in principle. Substitute the painting for a hologram if this seems more appropriate.

seems very reasonable. This point retains its force even after any "extra inner-object"[9] notion of experience is rejected so long as we take conscious experience (as with the painting) to be a *real* part of the process, regardless of its relationship with the rest of reality, and regardless of the veridical status of any particular conscious representation.

Thomasson says that "since there is also a phenomenology, I can turn my attention to that" and examine it (Thomasson, 2000, p.205). This is perfectly in keeping with the view that we are always at least *non-focally* conscious of our phenomenology; it's "there" (that is, it is present in experience) and so we can turn our attention to it with ease. But this line of thought sits at odds with the idea that we were, until introspecting, entirely *unconscious* of our phenomenology. On this view we would not just be shifting our attention to a particular aspect of our conscious content, we would be bringing into our conscious apprehension something that previously wasn't (phenomenally) *there* for us at all. This doesn't seem to describe the phenomenological data of introspection accurately: when we attend explicitly to the intrinsic quality of our experience it does not strike us as something that wasn't already there for us. It seems that we were in fact already conscious of it, even though we were not previously explicitly attending to it. Succinctly put, for Thomasson's view to work it needs to tackle The Novelty Problem. Claiming that the phenomenology is conscious or "present" prior to introspection even though its subject is unconscious of it does not adequately do this: the kind of novelty relevant here has to do with what the subject is conscious *of*. The qualitative character, claims Thomasson, has moved from something the subject is unconscious of to something the subject is conscious of. The Novelty Problem is therefore left untackled, and the introduction of a new category of consciousness (wherein something can be part of a subject's phenomenal consciousness but not something the subject is conscious of) seems questionable.

Ultimately, Thomasson needs qualitative character to be something the subject is conscious of in order for it to do the work afforded it by her theory. This is, at least, a plausible interpretation of the situation. Given that she holds phenomenology to be responsible for all of a subject's experiential contact with the world, she is thus committed to a UIA view. This entailment follows from two key premises: firstly, that

[9] i.e. the point of the painting metaphor applies even though our visual experiences, for example, are not little "inner pictures" that we look at.

experience is real (*really* real) and, secondly, that phenomenology is what makes anything seem anyway at all to a subject. If one holds both these premises to be true (and it seems very reasonable to do so) then one is committed to the ubiquity of inner awareness thesis due to the straightforward observation that phenomenology (experiential what-it's-likeness) cannot *be* without being something the subject is conscious of. It might be argued that assuming the truth of this observation is question-begging. The following response seems reasonable: it is analytical that phenomenology is itself always phenomenal. Being phenomenal, in this sense, reasonably means appearing to, or for, a subject, i.e. being something the subject is conscious of. To be phenomenal and yet not appear in this way is a highly counterintuitive idea (at odds with the standard employment of the concept), and so long as we can make good sense of an alternative, ubiquitous inner awareness for one, we should focus our efforts there. Given that there are good reasons to accept some kind of UIA view (as we shall see) and given that such a view does not require us to use the concept of "phenomenal" in such a counter-intuitive and non-standard way, we have little reason to accept middle-ground views of this kind as favourable.

§1.4 Affordance and The Novelty Problem

To take stock then: middle-ground views that hold phenomenology to be conscious face a difficulty when they deny that the subject of said phenomenology is conscious of it. They are forced towards a UIA view in order to account for the work done by qualitative character. There is, however, another (somewhat ingenious) option that a middle-grounder can take put forward by Tom McClelland (2014). On this view the subject is not conscious of the non-introspected experience (as with Thomasson), but is phenomenally conscious of the experience's *potential* to be introspected. This seems to go some way to addressing The Novelty Problem as although the experience itself is not pre-reflectively present to the subject, *phenomenal* traces of it are. The crucial difference with this view, as compared with more standard middle-ground views, is that our knowledge of our own phenomenology is not *entirely* absent from our conscious apprehension: our experience is, one might say, vicariously (and, crucially, *phenomenally*) present for us by way of its affording introspection.

McClelland draws our attention, firstly, to affordance as a general feature of conscious experience:

> If we wish to give a rich and accurate description of our phenomenology, we need to have the notion of affordances in our conceptual toolbox. When we perceive the world we are not merely passive spectators but rather active participants. Our potential to engage with our environment figures in our perceptual experience. The ball is not just given to us as red and round, it is given to us as *kickable*. (McClelland, 2014, p.16, original italics)

This seems to be the right thing to say. Further to this:

> Our opportunity to perform these acts figures in our experience: there is a manifest phenomenological difference between just seeing the ball and seeing it *as kickable*. (McClelland, 2014, p.16, original italics)

Affordance is not simply a matter of an unconscious desire or belief affecting our behaviour (though this may well be part of the whole story): it is phenomenally present for the subject or, in other words, the subject is conscious of it.[10] The suggestion is that something similar is happening with respect to our experiences themselves:

> Armed with the concept of affordances, we can offer a novel account of how inner awareness figures in our ordinary non-introspective experiences. Introspection is an action. All conscious states - or at least all ordinary conscious states of normal adult humans - are introspectable … My suggestion is that this ever-present potential for introspection actually figures in our experience. Your capacity to gain inner awareness of your concurrent conscious state colours *what it is like to be* in that state for you. Although our outer awareness of the world is not generally accompanied by an inner awareness of that very state, it is accompanied by an awareness of the *opportunity* for

[10] Although perhaps less obvious than other elements of our experience, the existence of this kind of mental phenomena has been recognised and defended for some time now: see Findlay (1955) for a particularly insightful investigation.

> introspection. In other words, *an affordance of introspectability is a ubiquitous feature of our phenomenology*. (McClelland, 2014, p.17, original italics)

McClelland's view concerns the relationship between consciousness and inner awareness generally, but it can also be employed as an interestingly novel approach to The Novelty Problem itself. Taken as such, however, it has its drawbacks. For one thing, as McClelland concedes, a reliance on potential introspectability in accounting for our (pre-reflective) inner awareness creates a distinction between normal human adults and any conscious beings incapable of introspection (perhaps certain other animals and human infants, depending on one's view of the mental abilities required for introspection). For some this distinction will be suspect and provide good reason to avoid explaining the UIA view as resulting from a misinterpretation of introspectability affordance. It is worth noting that we may not want to deny that affordance of this kind is part of a normal human adult's phenomenology - only that it is not *all there is* to (pre-reflective) inner awareness.

Further to this McClelland admits of the possibility that "in normal humans the capacity for introspection might go 'offline' during abnormal states of consciousness such as dreaming" (McClelland, 2014, p.18). Once again this introduces a distinction that we might be best avoiding - this time between what it is like to be awake and what it is like to be dreaming. There are, of course, a number of interesting differences between these two kinds of states (especially for those fortunate enough to experience lucid dreaming), but it is not obvious that there is such a global phenomenological difference as the one entailed by the affordance theory. Given that, as McClelland himself states, the affordance of introspection is meant to be a *ubiquitous* feature of waking experience this implies a dramatic difference in the nature of the phenomenology that it is absent from. But as strange as our experiences during sleep might be, they are still very much *experiences* in the fullest sense of the word: they are fundamentally the same kind of things we live through in our waking life. They are not "dreamed experiences" if what is meant by this is anything other than "real experiences during sleep". Lacking such a ubiquitous feature of normal experience, we should expect to see a *significant* difference in the nature of our phenomenology when dreaming and it is not obvious that we do.

Perhaps the affordance theorist can bite the bullet on these points though: infants and non-human animals may well be drastically different from us

phenomenologically speaking,[11] and perhaps we have no good reason to think otherwise aside from an intuition (and a fondness for our pets).[12] They might go on to point out that dreaming is indeed radically different from waking life: perhaps the affordance theory sheds light on one of the underlying reasons for this contrast. There is, however, a further problem confronting the affordance theorist.

The difficulty stems, once again, from the attempt to occupy a theoretical middle-ground between UIA and RT views, and comes to the fore as a result of the following question. What is it, exactly, that affords introspection? Everyday affordances seem to be tied to objects in some sense, as McClelland readily accepts:

> When kicking is afforded, for example, there must be a specific object that seems kickable to us. We never have a free-floating sense of *kickability* detached from any particular apparent object. (McClelland, 2014, p.19, original italics)

The trouble is that McClelland does not want to claim that any particular object affords introspection. It seems to make little sense to say, for example, that we can introspect a tree. One way out of this problem would be to adopt a RT view and say that we are never directly conscious of the intrinsic nature of our experiences: when we introspect an experience of a tree we just focus on certain parts of the tree's objective qualities, never getting to the intrinsic nature of the experience itself. But McClelland, rightly, wants to allow for our ability to be directly conscious of our own experiences. At the same time, as a middle-grounder, he wants to avoid becoming committed to a ubiquitous inner awareness. For this reason he cannot hold that it is our experiences that afford introspection, as this would entail a constant awareness of our own phenomenology (given that he holds the affordance of introspectability to be ubiquitous). He is, as a result, forced to claim that what affords introspection is our worldly situation. To be clear: the claim is that although the situation affords introspection, it is the experience that is introspected. To clarify:

[11] Peter Carruthers (1989), for example, has argued that the mental states of animals are nonconscious.

[12] The affordance theorist can also claim, as McClelland does (McClelland, 2014, p.19), that an affordance need not require any sophisticated judgement on the subject's behalf, and that therefore this kind of objection is misplaced.

> When you experience an affordance to dance in a nightclub, it is the *situation* - the music, the lights, the right degree of intoxication - that presents an opportunity for dancing. My suggestion is that introspection is like dancing in this respect. We can be aware of the opportunity to introspect in our current situation without having to be aware of any object *as a thing to be introspected*. Unlike with dancing, *every* ordinary situation presents an opportunity for introspection which is why the affordance of introspectability is a ubiquitous feature of our phenomenology. (McClelland, 2014, p.19, original italics)

McClelland adds that this picture of things is ultimately "answerable to [our] phenomenological reflection" (McClelland, 2014, p.20) and an idea that might take some mulling over before striking one as convincing. That said it does not seem, on the face of it, to be obviously wrong. Furthermore, it has a substantial advantage over the standard middle-ground views: its concept of "introspectability affordance" acknowledges the phenomenal presence of experience to its subject in a much fuller sense than "unconscious knowledge" and "availability for reflection" can (although, of course, not in the fullest sense we see granted by the concept of ubiquitous inner awareness). However, once again, The Novelty Problem comes into view.

On the affordance view what we are conscious of prior to introspection are, perhaps among other things, worldly objects and our worldly situation. McClelland does not give a full catalogue of the kinds of things we are typically conscious of, but the relevant point is that we are not conscious of our experience. Until introspection occurs this is very much in keeping with the previously discussed transparency views: although our experience is not something we are conscious of it allows us to be conscious of, roughly speaking, the world. For McClelland the situation affords introspection, the concurrent experience is introspected and we thus become conscious of our experience. It is here that the novelty objection reappears: if we had moved from a state of being conscious of the situation to being conscious of experience *itself* we should expect to notice some kind of substantial phenomenological novelty.

This point assumes (as I take it McClelland does) that experiential qualitative character is real. The affordance theory claims that a subject can move from an awareness of everyday objects and one's situation to an awareness of the stuff that strikes us as being *so at odds* with these worldly objects that it has, for many, required

an ontological category all of its own (a category in part defined as *the opposite* of a worldly-object category), and yet not notice any major phenomenal difference. It is important to note that even if some form of reductive materialism were true it would still be the case that the experiential-qualitative dimension of our lives *strikes* us as being substantially different in nature from ordinary worldly objects. Very few people are inclined to think, for example, that there is anything it is like for a tree to be a tree or for an electron to be an electron: such objects strike us as fundamentally different kinds of things when compared to experiences (precisely in their seeming lack of any experientiality of their own).[13] It is of course true that there are differences between our everyday awareness of the world and our introspective experience: there is a change in the focus of our attention and quite often (although not necessarily) a reconceptualisation (such as when we consider our current experience *as* an experience). But these changes all occur against the background of a general familiarity that we should expect to find disrupted if we really did bring experientiality into our conscious apprehension only upon introspection. This general point assumes that we take both experience and our current scientifically informed picture of reality seriously, in the way outlined previously. Direct realists arguably fail to do the latter, and as such the arguments presented here are not aimed at them.

In order to avoid The Novelty Problem McClelland could claim that, even upon introspection, we are in fact still only conscious of worldly objects, properties and situations, but that we have in some sense reconceptualised them. The affordance theory would at this point, however, have collapsed into a kind of RT view; we would not be directly conscious of the intrinsic nature of our experiences, but instead would be thinking about worldly-objects using a different conceptual picture. It is tempting to see middle-ground views as trying to have their cake and eat it too: in attempting to maintain phenomenological accuracy (i.e. by accepting our ability to be directly conscious of our experiences) whilst at the same time wishing to deny the ubiquity of inner awareness, they struggle to adequately deal with the The Novelty Problem.

Aspects of the middle-ground views covered here can indeed be appropriated into a workable theory that avoids the novelty objection. It seems, however, that in order to do so they must move towards RT or UIA views. Both of these positions have no

[13] Panpsychists are all too familiar with this natural impression, and have to work hard to convince their opponents to part with it when proposing that experientiality is a fundamental part of objective reality.

problem with the lack of phenomenal novelty we find upon introspection: on one view we are always already conscious of the intrinsic qualities of our experience, on the other view we never are. Whilst it is true that *both* of these views can be made to appear quite counterintuitive, depending on the details supplied, there are reasons to hold that the notion of radical transparency is the least promising of the two. Firstly, there are a number of strong arguments in favour of ubiquitous inner awareness and, secondly, radical transparency faces serious difficulties in its own right.

§1.5 Radical transparency and phenomenal novelty

One of the chief problems confronting RT views is related to the notion of phenomenal novelty already discussed. As we saw, the lack of substantial phenomenal novelty found during introspection gains its dialectical power from the contrasting ways in which the experiential-qualitative aspect of reality and ordinary physical things (lacking any experientiality of their own) impress themselves upon us (regardless of the *actual* relationship between the two). Their apparent incommensurability has led to a number of different versions of dualism and idealism, and a recent surge of interest in panpsychism. It is also precisely why eliminative materialists feel the need to *eliminate* our experiences (as we typically know and describe them): they don't seem to fit with our standard picture of wordly objects. But if RT is true, if we are never conscious of the intrinsic qualities of our experience - "period" (Tye, 2014, p.40, original underlining) - then why is it that we have such a powerful "intuition of distinctness" (Papineau, 2002, p.3)? Why did this ontological category, defined by its opposition to the everyday world of objects, get off the ground in the first place?

To claim that the motivation for such a category only comes about as a result of deep (and mistaken) philosophising ignores how readily we are apt (prior, even, to any consideration of ontology) to recognise the apparent distinction at issue. It is not unreasonable to speculate that one of the most commonly held beliefs concerning selfhood in human culture involves a presumed distinction between the body and its "soul". Intuitively regarding experiences and worldly-objects as different kinds of things would go some way to explaining this (and to claim that the motivation might come from prior religious commitments is likely getting things the wrong way round).

RT defenders lack an adequate explanation for this state of affairs and it is difficult to see what this might be on their account, given that they deny our ability to be conscious of the intrinsic qualities of our experiences at all. Once again, UIA defenders have, on the other hand, a ready explanation: we are prone to make such a distinction because we are indeed both directly conscious of the experiential-qualitative aspect of reality and at the same time conscious of worldly-objects, and they (accurately or not) strike us as fundamentally different features of reality.

A further serious problem confronting radical transparency has already been hinted at: the guiding impetus for one of its central claims is based on mistakenly assuming that there is only one model of introspection available to those wishing to defend inner awareness. This is the view that introspecting an experience is akin to shifting the focus of our attention away from worldly objects and "inwards" towards a *further* set of internal objects. However, if a different take on inner awareness can be successfully developed that does not require such a mistaken view of introspection, then one of the chief motivating reasons for accepting radical transparency disappears. In other words, if we can make good sense of inner awareness without having to deny something as basic as our ability to be directly conscious of the intrinsic qualities of our experiences, then this is preferable. (Middle-ground views are examples of just such an attempt but, as we have seen, face various problems.) On the assumption that there is a workable alternative view, the outlandish nature of RT counts against it.

No knock-down argument against either RT or middle-ground views has been offered: there are possible responses that we have not dealt with, and perhaps responses that are in fact yet to be articulated. Having said as much, the kinds of issues we have seen serve to highlight the *ease* with which UIA deals with the lack of novelty we find upon introspection. Taken in tandem with the forthcoming arguments in §2, this adds to the plausibility of such a view. We will, then, accept the existence of inner awareness and hold that there are strong reasons to think that it is a ubiquitous feature of experience. What is now needed, then, is a solid *general* account of inner awareness. If it is true that, in being conscious, we are always already conscious of our experience (i.e. we are conscious of our phenomenal consciousness) then we need a convincing account of the nature of this awareness. As we will come to see, the most promising answers to this question (when combined with a very reasonable account of temporal consciousness) entail some surprising things about the

relationship between consciousness and selfhood, resulting in a radical view concerning our very own nature.

2: Higher-order Theories and Inner Awareness

§2.1 The landscape

What we require now, then, is a convincing account of the nature of this inner awareness. Let's put aside its *ubiquity* for now and consider a different problem. Granting that we can, at least sometimes, be directly aware of our own experiences the following question can be posed: what is the relationship between an experience and our awareness of it? The ensuing debate can be framed in the following way: is a subject's awareness of any given experience somehow contained within that very experience, or is it a secondary awareness directed at the original (and *distinct*) conscious state? We can label the first picture the "intrinsic-inner awareness view" and the second the "higher-order view".

The foundational maxim upon which the various higher-order theories are built can be put the following way:

> When a mental state is conscious, one is to some degree and in some way conscious of that state. (Rosenthal, 1994, p.356)

This is a highly intuitive starting point. Consider the occurrence of a typical conscious mental state such as a painful sensation. We care to be rid of such a state precisely because we are conscious of it (and, usually, dislike it). To say otherwise is to employ a radically non-standard and counterintuitive notion of consciousness as we saw with Thomasson in §1.3 (pp.26-31). Consider being told by some sufficiently advanced future neuroscientist that although you are currently not conscious of any pain, one of your present conscious mental states is an agonisingly painful sensation. It is a conscious state, you are assured, but not one you are conscious of. Given that you feel quite fine in fact (i.e. you are not conscious of any pain), it seems something has gone wrong with the diagnosis (or the conceptual framework within which the diagnosis was formed). Intuitively, what we mean by a "conscious mental state" is one that we are *conscious of*. Recall a memory from your childhood: a moment ago this was an unconscious mental state of yours, and now it is a conscious one. What accounts for

its changing status? An eminently reasonable (and, some might say, trivially obvious) answer is Rosenthal's: it has now become a state you are conscious of. With the previous issue of novelty we saw that both sides of the debate shared a highly intuitive fundamental claim (namely that there is a lack of phenomenal novelty when introspecting), and it was argued that this agreement, considered in the context of such markedly contrasting theories, counted in its favour. We see a similar development in the coming discussion in that its competing theories give fundamentally different accounts of inner awareness, yet all defend the common-sense notion that a conscious state is one that its subject is conscious of. Once again, in the midst of such different agendas, the sharing of this claim stands in its favour.

There are a number of different ways higher-order theories have attempted to account for this inner awareness, but their unifying claim is that a subject's awareness of its own conscious state is of a higher-order than the original target state (which is to say; it is a further and *distinct* state or act of consciousness). When first confronted with this general picture of inner awareness it can seem very reasonable indeed. When we are conscious of some worldly object before us it is (in part) because that object is taken as an intentional object of some conscious state of ours. To be conscious of something "external" is to have a conscious state directed at it. It seems to follow, then, that we should say something similar regarding inner awareness: when we are conscious of some conscious state of ours it is because that state is an intentional object of some further conscious state of ours. Questions can then be asked as to the precise nature of this secondary awareness (whether it is perception-like or thought-based, for example). Unsurprisingly, however, things are not quite so simple. A number of damaging arguments against higher-order theories can be made. It seems that each variation of the higher-order take on inner awareness brings with it more problems that it solves, and that consequently a different kind of theory is called for. Before turning to just such a view, it will be instructive to see how and why the higher-order theories run into trouble.

§2.2 Attention and consciousness

This is Armstrong's view of inner awareness:

> I think it is an additional form of perception, or, a little more cautiously, it is something that resembles perception … a perception-like awareness of current states and activities in our own mind. (Armstrong, 1981, p.724)

Again, the rationale for adopting such a perspective seems, on the face of it, quite reasonable: we are conscious of objects in the world around us due to acts of perception. It seems to follow that, similarly, when we are aware of our own mental states it must be because we have some kind of perceptual (or perceptual-like) awareness of them. Armstrong pushes the analogy even further: just as external objects can exist unperceived, so too can conscious states. This is partly why Armstrong takes awareness to be distinct from the conscious mental states it takes as objects: they can exist even in its absence. He claims that most of us are in fact quite familiar with such a state of affairs, and puts forward the following example:

> After driving for long periods of time, particularly at night, it is possible to "come to" and realize that for some time past one has been driving without being aware of what one has been doing. (Armstrong, 1981, p.723)

There are a number of problems with Armstrong's view in virtue of its being a higher-order theory, but before delving into the specifics it will be useful to challenge this seemingly innocent phenomenological description. The kind of experience Armstrong is recalling is, although not typical, one that many of us can no doubt relate to. Upon "coming to" (as Armstrong knowingly calls it) there can indeed be a sense of surprise at having navigated for such a stretch of time without having paid much attention to the task at hand. But paying a small amount of attention is not the same as paying no attention at all. More importantly, even if we grant that zero attention was being paid to the acts of perceiving and navigating the road (the driver being, in the fullest sense, on "autopilot") this does not secure the conclusion that the driver was unconscious of such things. To achieve this one would have to establish a necessary connection between attention and consciousness, with the latter dependant on the former. This seems an implausible project when we pause to consider our own typical conscious streams fully, particularly the virtually constant "phenomenal background" to which we rarely attend:

> [There can be] experience without conscious awareness, in the sense of 'awareness' as attention or recognition. But experience that is not attended to is still experience. Indeed, it is plausible to suppose that the bulk of our consciousness consists of this sort of unnoticed experience. I will call this sphere of experience the *phenomenal background*. The phenomenal background goes largely unnoticed because it is constantly present for as long as we are awake (and often while we dream). Most experiences that go on long enough for us to become habituated to them (but which do not cease altogether) will sink into this background, for example the sound of a refrigerator, or the noise of a car engine. (Dainton, 2006, p.31, original italics)

The humming of an electrical appliance can be a particularly vivid example of the phenomenal background: consider the following phenomenological description in response to Armstrong's autopilot story.

You are reading in a quiet room when you notice a subtle but constant humming coming from a nearby electrical appliance, perhaps the light above you or your personal computer nearby. It may have gone unnoticed during previous reading sessions in this room but your attention has been called to it now (perhaps because you are reading about background humming). You endeavour to ignore it. You succeed and become fully immersed in your reading once again. Half an hour passes when suddenly the noise stops. You are immediately aware of its departure and in fact welcome the contrasting peace. Though you had not been attending to the sound just prior to its cessation (you were deeply engaged in thought), the difference in your experience is concretely tangible and the change in its quality is *directly experienced* (not merely inferred). It was not as if you noticed some kind of change, examined the contents of your experience for a clue as to its nature and then, remembering your previous experience, deduced that it was the noise stopping. This picture overcomplicates the matter: it is more accurate to say that you simply *heard the noise stopping*. Which is to say that you experienced the transition from humming to (relative) silence. For this to occur the humming must have been part of your conscious experience prior to its cessation, even though you were not explicitly attending to it (or else the transition would not have been *directly* experienced). In short, it was in the phenomenal background (along with much else).

The phenomenal background is rich and multi-faceted but for the present point it is enough to focus on just one part of it, namely its "world-presenting perceptual experience, what we see, hear, touch, smell and taste" (Dainton, 2006, p.31). Dainton goes on:

> The content of this experience is nothing less than the surrounding world: the ground underfoot, rooms, walls and furniture, streets, fields and trees, animals and people, the sky above - these are all parts of the phenomenal background, they all feature in our experience, for the most part unnoticed, as we go about our ordinary business. (Dainton, 2006, p.31)

Staying only with this aspect of the phenomenal background for now, we can say that the relevant features of the world surrounding a deeply distracted driver (the road, the other cars, the twists and turns, the dashboard, the feel of the wheel etc.) have receded into the phenomenal background (or, thinking in less binary terms, have at least moved towards it) when little or no attention is being paid to them. But in that case there is still something it is like, for example, to be seeing the road in front of you - even if this is not at the forefront of your mind. The road and the scene before you is not a *phenomenal void* (such as the "visual space" outside of your periphery and behind your head right now). The claim that we are only ever conscious of what we are explicitly attending to leaves out *so much* of the typical moment of human conscious experience.

Suppose your attention is now called to the many elements of your experience that are currently in your phenomenal background (necessarily, of course, therefore bringing them into the foreground in the process). Consider the feel of the chair you are sitting on, the distant noises of traffic or people or birds, the touch of the clothes you are wearing, the feeling of your feet being enclosed by your shoes, your general mood, the vague shapes and patches of colour in your peripheral vision. This is a very small list compared to the length a full catalogue of an average conscious moment would need to be. Does it really seem reasonable to suppose that before your attention was called to these aspects of your experience they were all *nothing at all* for you, experientially speaking? Whilst there doesn't seem to be anything self-contradictory in this idea, we would be better off if we can accept inner awareness without

becoming committed to such an implausibly impoverished view of what it is typically like to be conscious.

Armstrong is well aware of the oddness of claiming that we are *unconscious* of conscious mental activity to which we are not attending, and tries to avoid saying as much. He identifies three kinds of consciousness: minimal, perceptual and introspective (Armstrong, 1981, p.723). He defines them as follows: when a person is dreaming they are only minimally conscious. If awake and perceiving the world they are both minimally and perceptually conscious. If they are attending to their mental activity then they are introspectively conscious (i.e. inner awareness is occurring). So for Armstrong the autopilot driver is minimally conscious (in virtue of having some kind of mental activity going on) and perceptually conscious (in virtue of not being asleep) but lacks introspective consciousness of the act(s) of driving. With this terminology in place Armstrong can avoid claiming that the driver is unconscious of the activity of driving and its relevant environmental features. But all he means by this is that, with regards to driving, "there is mental activity going on" (Armstrong, 1981, p.723) (minimal consciousness is occurring) and the driver is not currently dreaming (perceptual consciousness is occurring). There is no room in this picture for any *phenomenal* consciousness (with regards to the relevant aspects of driving). On this view there is nothing it is like for the driver to be negotiating the twists and turns in the road, to be seeing the other vehicles or to be handling the wheel. All of this is *entirely absent* from the currently occurring experiential reality of the driver. There may be good reasons to employ the concept "consciousness" as Armstrong does, but the claim that *all* unattended mental activity is experientially non-existent for its subject is hard to take seriously. Once again, a view of inner awareness that doesn't make such implausible claims would be preferable.

Suppose that the driver, upon "coming-to", attempts to settle this philosophical dispute by recalling the past five minutes of her autopilot journey. Suppose she is able to remember some event along the way (a turn in the road or a pedestrian she gave way to) and in fact remembers this moment in just the usual way: as an experience of the very same kind as any other. That this could occur seems highly likely (the reader is invited to investigate this claim for themselves the next time they "come-to" during some routine activity). Armstrong would now need to account for how this moment of mental activity could be recalled in such a *typical* way given that it was not originally part of any experience. Is the phenomenality mysteriously (and mistakenly) projected

backwards? He might claim that it is precisely *because* the driver is able to recall what it was like to undergo that particular mental activity that it must have been introspectively conscious in the first place. Perhaps most of the autopilot period was introspectively unconscious, save for this particular moment (and maybe a few others). This would seem to be somewhat of an ad hoc response.[14] Furthermore, it would admit of the following: that (at least sometimes) someone can feel justified in claiming that they were, in Armstrong's terminology, not "introspectively conscious" of some mental activity when in fact they were. This would count against Armstrong in that his evidence for supposing a lack of phenomenal consciousness (i.e. that one has a sense of "coming-to" and cannot remember the relevant experience(s)) is, at least sometimes, unreliable. Contrastingly, the notion of the phenomenal background comes away unscathed and, in fact, neatly explains why someone might mistakenly *assume* a lack of (full-blown) consciousness of any given activity (simply put: it was not the focus of their attention at that time and thus more difficult (perhaps even impossible) to recall).

As a final point on the phenomenological accuracy of Armstrong's autopilot example, consider the *substantial* experiential difference that exists between dreamless sleep and conscious waking life. These are not typically taken to constitute two contrasting instances of the same general kind of thing. For many, they are in part defined by the very absence of the other:

> Consciousness consists of inner, qualitative, subjective states and processes of sentience or awareness. Consciousness, so defined, begins when we wake in the morning from a dreamless sleep and continues until we fall asleep again, die, go into a coma, or otherwise become "unconscious". (Searle, 2000, p.559)

When we are dreamlessly asleep we are not experiencing at all (or, at least, this seems a very reasonable position to hold on the issue - one we shall return to in §5.5). Armstrong claims that although the (driving-relevant) mental activity is "perceptually" conscious it lacks *any* experiential component. The problem here is

[14] This would be particularly problematic if a third party were to *randomly* prompt experience-recall of this kind (by asking questions such as "Do you remember the bend in the road by the post office?"). If the *successful* recall frequency was similar to that of typical (non-absent-minded) stretches of experience then Armstrong would have some explaining to do. Unfortunately, this kind of data would be incredibly difficult to obtain.

that although there is indeed a qualitative difference between attentive and inattentive conscious mental activity, the contrast is not as dramatic as Armstrong's theory needs it to be (where the latter is claimed to have no subjective quality *at all*). Dainton argues the following on this very point:

> Armstrong's day-dreaming driver may not have been paying attention to what he could see on the road ahead, he may have been instantly forgetting what he was seeing, but this does not mean that he wasn't consciously seeing anything at all during this period. To appreciate this fact it suffices to imagine how different his overall state of consciousness would be in the absence of all visual, auditory and bodily experience: the experience of driving on auto-pilot is certainly different from the experience of driving with one's attention fully focused on the job in hand, but it is nothing like driving in total darkness and silence - or being dreamlessly asleep! (Dainton, 2004a, p.8)

§2.3 Inattentional blindness

Relatedly, the notion of "inattentional blindness" has been used to defend the claim that attention is necessary for consciousness. One of the more striking psychological experiments dealing with this was conducted by Simons and Chabris (1999). A version of the video used in the experiment is easily searchable online and is worth watching before reading on (though, of course, its effects might be lessened for you given the information already supplied about its context). Simons and Chabris begin their account by identifying the phenomenon under examination:

> Perhaps you have had the following experience: you are searching for an open seat in a crowded movie theater. After scanning for several minutes, you eventually spot one and sit down. The next day, your friends ask why you ignored them at the theater. They were waving at you, and you looked right at them but did not see them. Just as we sometimes overlook our friends in a crowded room, we occasionally fail to notice changes to the appearance of those around us. We have all had the embarrassing experience of failing to notice when a friend or colleague shaves off a beard, gets a haircut, or starts wearing

> contact lenses. We feel that we perceive and remember everything around us, and we take the occasional blindness to visual details to be an unusual exception. (Simons and Chabris, 1999, p.1059)

Just how accurate they are in their final claim is questionable, but the general experience they describe is certainly an easily recognisable one. Such inattentional blindness is a powerful tool for certain professions: pickpockets and magicians employ exactly this when they divert the attention of their victim or audience. A person's failure to notice a substantial change in their immediate environment can also provoke a comical effect: in a section from a popular television show a prankster approaches a member of the public in the street and asks for directions. As the helpful stranger is looking at the map held by the prankster an accomplice smoothly takes his place. So long as this is done whilst the helper's attention is on the map, the change goes unnoticed even when eye contact is subsequently re-established. The extent of the comical effect increases in conjunction with the extremity of the contrast between the prankster and his accomplice. That such occurrences can be found amusing is presumably linked to the surprise we feel in witnessing someone fail to notice something so substantial occurring in their midst: although we may not feel that we perceive and remember "everything around us", as was claimed above, we do feel that we perceive most things of significance in our immediate environment.

The experiment conducted by Simons and Chabris (1999) is similarly amusing: participants were asked to watch a video of a group of basketball players and count the number of times a ball is passed between them. In one version of the video a woman wearing a full-body gorilla suit enters the frame and leaves after five seconds, and in another version a woman enters holding an umbrella. After watching the video the participants were asked to provide answers to a number of questions, one of which concerned whether or not they had seen anything unusual in the footage. The results were interesting and are effectively summarised by Daniel J. Simons as follows:

> On average, approximately 35% of subjects did not see the fully visible umbrella woman and gorilla. In one extra condition, the … gorilla stopped halfway across the display, turned to face the camera, thumped its chest, and then exited on the other side of the screen … half of the observers did not see it! In fact, when we showed the video again after explaining what had occurred,

> observers were often shocked, sometimes even exclaiming, 'I missed *that*?!' Most observers intuitively believe that unusual events will explicitly capture attention. (Simons, 2000, p.152)

To avoid question begging the above claim should in fact be that approximately 35% of subjects did not *report* having seen anything unusual. This speaks to an alternative interpretation of the results. It is possible that the subjects may have in fact seen the gorilla only to instantly forget it:

> Subjects might attend to an object, consciously perceive it, and then forget it by the time they are asked about it. Although we can safely argue that a stimulus was attended to if subjects can recall it, we cannot necessarily infer that it was unattended to if it was not recalled. In practice, these two explanations, blindness and amnesia, might be empirically inseparable. No matter how quickly subjects can be asked about a critical event, the questioning will still occur after the event. If observers fail to report it, proponents of the inattentional amnesia hypothesis can claim a failure of memory rather than a failure of perception. (Simons, 2000, p.153)

It may at first appear unlikely that one could consciously perceive such a strange thing and then forget it. This concern becomes less pressing when we consider the different ways in which an object can be consciously perceived. One such way is for the object to make a phenomenal difference to the overall qualitative character of the experience and *also* be conceptualised, i.e. consciously singled out from the other objects perceived and explicitly recognised *as* the particular kind of object it is. As we saw with our discussion of the phenomenal background, however, this is not the only way that an entity can be consciously perceived: the feel of the chair on your body is something that you were conscious of even before this sentence urged you to attend to it - it was making a phenomenological difference to the overall character of your experience. The case of the gorilla might well be explainable in the same way: its visual presence was making a phenomenological difference to the overall character of the experience, but it was not explicitly attended to and thus did not reach the level of conceptualisation. Having failed to become conceptualised it is not surprising that it would also fail to be stored in accessible memory. It may even be that it *was*

successfully conceptualised as a gorilla (or, at any rate, as a person in a gorilla suit) but not attended to. In this case it *is* seen as a gorilla figure but not explicitly *recognised* as such: it is not *re*-cognised to the extent necessary for it to enter into explicit attention or memory - it flits in and out of conscious awareness before such a process can occur. On either view, you - as the observer - were conscious of the gorilla figure. Although the object was *in front* of you it was part of your phenomenal *back*ground: it was part of the vast array of phenomenal content that, although unattended to, goes to make up the overall qualitative character of any given stretch of your experience. As Simons notes (Simons, 2000, p153), this alternative explanation accounts for the experimental data just as well as the blindness hypothesis. It seems, then, that such considerations alone do not force our acceptance of attention as a prerequisite for conscious experience.

§2.4 The phenomenal background in focus

Watzl (2011) has proposed a theory of attention that goes some way to making the notion of the phenomenal background and its claims concerning attention even more attractive. A core claim in his argument is the following: "*consciously attending to something in part consists in consciously experiencing what is unattended in characteristic peripheral ways*" (Watzl, 2011, p.155, original italics). To clarify:

> When one attends to something, other aspects of one's experience recede to the periphery. This affects their phenomenology, just as it affects the experience of what your attention is focused on. Instead of asking about the phenomenology of attention, we might ask about the phenomenology of the periphery. Focus and periphery seem to be two sides of the same coin. (Watzl, 2011, p.156)

He presents the following example: suppose that you are enjoying a musical performance by a jazz band whilst at the same time undergoing a pain in your left foot. Suppose that you decide to try and ignore the pain by closing your eyes and concentrating your attention on the sound of the saxophone. Your attention is now focused explicitly on the saxophone but, so Watzl argues, there is more to the overall qualitative character of your experience than just that: you do not only experience the

saxophone when you are attending to it - you also experience the piano and the pain as being in the periphery (Watzl, 2011, p.156). Crucially, Watzl claims that this background is *structured*: when you attend to the saxophone the piano is "experienced as relevant for or close to the experience of the melody played by the saxophone" (Watzl, 2011, p.156). Your pain, irrelevant as it is to your desire to hear the music, might be further towards the "fringe" owing to the merely *marginal* awareness you have of it (depending, of course, on its severity) (Watzl, 2011, p.156). Watzl references a similarly persuasive case made by Sartre:

> When I enter this café to search for Pierre, there is formed a synthetic organization of all the objects in the café, on the ground of which Pierre is given as about to appear ... [If] I should finally discover Pierre, my intuition would be filled by a solid element, I should be suddenly arrested by his face and the whole café would organize itself around him as a discrete presence. (Sartre, 2010, pp.33-4)

The accuracy of this description will be obvious to many. Its reasonableness is especially bolstered when we consider the proposed phenomenology of the opposing claim that unattended content is unconscious. In the case of the café scenario, recognising a friend's face and attending explicitly to it should result in the momentary annihilation of any phenomenal content to which one is not attending. Suppose that you had been attending to the sounds in your environment. On this account the sounds of muffled conversations and clinking cutlery will suddenly vanish from your phenomenology the moment you recognise and attend to your friend's face. It is not at all clear that this is an accurate description of what it would be like to undergo such a meeting. Sartre's model seems far more accurate: the unattended content "organises" itself around your friend's face. In other words, it moves into the phenomenal background. For Watzl, this is because "consciously attending to something consists in the conscious mental process of structuring one's stream of consciousness so that some parts of it are more central than others" (Watzl, 2011, p.158). Watzl argues that the structure of this stream is determined by "attentional relations" between its parts: the simplest kind of relation is that of some content (x) being peripheral to some other content (y) (Watzl, 2011, p.158). Crucially, these relations are *phenomenal* and thus make a difference to the overall qualitative

character of the experience. Remain visually focused on the '+' below whilst consciously attending to, firstly, the left '§'' and then the right one:

§ + §

The qualitative character of your experience changes depending on which '§' you are explicitly attending to, even though it is clear that both symbols remain part of your phenomenology throughout the switching process. What it is like for you to focus on the left one is different from what it is like for you to focus on the right one. This difference cannot be accounted for by the phenomenal absence of the unattended content, as it is clearly still present. What then *is* the best way to account for the phenomenological difference? Watzl has a convincing model: the structural relationship between the central content and the peripheral content is changing when you switch focus. When you are attending to the left '§' the right one is experienced as being peripheral to the left one and vice versa upon switching:

> The relevant structure has as its primitive the phenomenal *peripherality* relation "*x* is peripheral to *y*" (importantly distinct from any *spatial* form of peripherality). Consider the case where you are focusing your attention only on the sound of the piano. In the corresponding attentional structure, all other parts of your experience are peripheral to your experience of that sound … Consciously attending to the piano consists in creating and sustaining a total state of consciousness where your experience of the piano is central to the other parts of your experience. We can now also make sense of the idea that in a scenario like this your experience of pain is likely to be *more* peripheral than your experience of the saxophone. Your pain experience is peripheral not only to your experience of the piano, but also to your experience of the saxophone. (Watzl, 2011, p.160, original italics)

As a phenomenological description of what it is like to attend to an aspect of one's phenomenology, Watzl's account is convincing. It goes without saying that you cannot experience some content *y* as peripheral to *x*, if you fail to be conscious of *y*. As such the claim that attention is a necessary condition for consciousness is suspect.

This issue concerning unattended content is difficult to settle with absolute certainty. One reason for this is the nature of the subject matter: usually when we intend to examine or investigate some issue we focus our explicit attention on it. This is, however, precisely what we *cannot* do when the matter at hand is unattended content: as soon as we focus on it directly it becomes attended to. This has prompted a further concern with regards to the reliability of claims relating to this kind of content, which has come to be known as the "refrigerator light fallacy":

> This is the fallacy, perhaps committed by a technologically naïve person, of thinking the light is always on in the refrigerator because whenever he opens the refrigerator the light is on. We can see how the fallacy applies to the domain of self-consciousness. We start with a point about what's immediately knowable, hence reportable, if asked. The query ("what are you doing?") is, metaphorically, the opening of the refrigerator. But just as it doesn't follow from the light being on when we open the refrigerator that the light is always on, so it doesn't follow from our being able to report knowingly on our conscious lives when asked that our conscious lives always includes self-consciousness. The fallacy is particularly inviting when we engage in phenomenological reflection. After all, to reflect on the structure and character of our own experience is an intensely self-conscious enterprise. As soon as we've set off on the investigation, we've "opened the refrigerator." Unsurprisingly, self-consciousness turns up wherever we look. (Schear, 2009, p.101)

Although Schear is making the above points in relation to inner awareness in general, they can also be levelled at those who claim that unattended content can be conscious. In a nutshell the argument is as follows: in attending to some content and finding it to be conscious, it does not follow that it was previously so. It may have been unconscious but still capable of pulling your attention in its direction. As we saw in §1.3 and §1.4 (pp.24-36), however, there are good reasons to believe that such a changing status in content, from unconscious to conscious, would result in phenomenological novelty of the kind that we do *not* find when engaged in such a practise. A ready explanation for this lack of novelty is that the peripheral content was indeed something you were already conscious of. Accepting that this is a particularly

tricky area to investigate does not mean accepting that we are incapable of making informed judgments on the issue. For one thing we can use our short-term memory: "[we] can 'replay', perhaps repeatedly, the past few moments of our experience, trying to remember it as best we can" (Dainton, 2006, p.33). We can attend to aspects of our experience that we failed to notice when we were undergoing them and can, as it were, retroactively attend to them. If, in so doing, you recall some content that you identify as having been unattended at the time then there are three possible explanations for this: either you have confabulated some content, or you have mistakenly remembered some content to which you were attending as unattended, or you have indeed correctly recalled some unattended content. It is difficult to see why we should hold the third option to be *impossible*. There is also another way we can gain access to the phenomenal background:

> We can make judgements about the phenomenal background more or less as it happens. Try the following experiment. Focus your attention as hard as you can onto the page in front of you, onto its colour or texture; keep your attention focused here, and while doing so describe out loud something else you can perceive, e.g. the colour of the walls that you can see in your peripheral vision, or any sounds you can hear. Suppose the walls are green: you can notice and report on this without significantly lessening the degree of attention you are paying to the page in front of you. There will probably be some reduction in the degree of attention you are paying to the page, but not a great deal. The important point is that you can register something of the character of the contents of your peripheral experience without focusing your attention onto your peripheral experience itself. We can call this procedure *passive introspection*. (Dainton, 2006, p.33, original italics)

Not everyone will find such techniques convincing but considered in conjunction with Watzl's account, Dainton's notion of the phenomenal background, and the problems facing Armstrong's autopilot scenario, they go towards the strong case that can be made for unattended conscious content. We have good reasons, then, to disagree with William James's well-known claim that "[*my*] *experience is what I agree to attend to"* (James, 1890, p.402, original italics). *This* motivation for adopting a higher-order view of inner awareness is not free from attack.

As was stated previously, there have been a number of arguments levelled against the higher-order views. There are also a number of ways to group these arguments. MacKenzie (2007) puts forward a categorisation that also takes into account the related debates from Indian Philosophy and thus provides a comprehensive overview to jump off from. Concentrating on the most powerful objections facing the higher-order views a triple-pronged attack can be made: it consists of what MacKenzie calls "the objectification argument", "the reportability argument" and "the regress argument" (MacKenzie, 2007). Armstrong's questionable phenomenology regarding the autopilot driver stems from issues concerning the objectification argument.

§2.5 The objectification argument

In a nutshell the objectification argument holds that a perceptual (or perceptual-like) model of inner awareness attempts to conceive of subjectivity in only objective terms, and that this is inappropriate (MacKenzie, 2007, p.58). More specifically: it misdesribes the way in which a subject is typically acquainted with its own subjectivity. Experiences, for example, are mistakenly characterised as perceptual objects, partly owing to the fact that "it is easy to lapse into talking as though experiences are things we perceive or observe, in essentially the same way as we perceive or observe ordinary physical things" (Dainton, 2006, p.44). Just why this mistaken characterisation of inner awareness is so seductive is easy to see: focus your attention on the feeling of your body sat in your seat. Now move your attention to your visual experience. Finally, move it to focus on the sounds in your environment. It can seem, on the face of it, that this kind of activity is a matter of looking around at different things: "[it] might seem as though [you] have a single sensory faculty, akin to an eye, a sensory organ of a special kind which [you] can point or focus wherever [you] like" (Dainton, 2006, p.45). Just as our eyes provide us with visual objects and our ears with the auditory variety, it can seem that a special organ of inner-perception is able to apprehend all manner of experiences as objects, regardless of their modality. But this picture of inner awareness is problematic: it entails a duplication of experiences that is neither parsimonious nor phenomenologically accurate:

> If we really did have this additional organ, and if it worked like our other sensory organs, our sensory experiences would be items that causally interacted with this additional organ to generate a *second* set of experiences - everything would be experienced twice over. Or rather, it would be more correct to say that the only experiences we would have - if we regard ourselves as the inner eye - would be experiences generated by the additional organ. After all, the experiences produced by our eyes and ears would be as absent from our consciousness as the physical things in the environment which causally stimulate these organs. These absurdities are the result of taking introspection to be a form of sensory perception on a par with ordinary sense perception. (Dainton, 2006, p.45, original italics)

A further possible reason for the allure of thinking this way might be traced to language. Janzen, in attempting to account for the popularity of this erroneous model, says just that:

> I would hazard a surmise that this is due, in large part, to the misleading surface structure of the grammar of perception sentences. Expressions like 'I had a perceptual experience' and 'I am having a perceptual experience', i.e., expressions in which the term 'perceptual experience' is grammatically the accusative object, are perfectly licit. But we should not be misled by the grammar of these expressions. To have a perceptual experience is simply to be in a conscious perceptual state; it is not to possess an object that happens to be a perceptual experience. (Janzen, 2006, p.60)

Having said as much, it would be rash to deny our ability to objectify our experiences in *any sense at all*. It is quite possible now, for example, for you to focus your attention on your experience of a particular object presently in your visual field. By focusing on what it is like for you to have that experience you are "objectifying" it in a *certain* sense: you are seemingly adopting the position of a spectator on an aspect of your experience. This is not at issue. The objectification argument against higher-order theories need not dispute *this* sense of "objectifying". What it does seek to show is that, firstly, when an experience is reflected on this act is not relevantly similar to sense perception, secondly; reflection is not in fact the typical way we are made aware

of our own experiences and, thirdly, that if subjectivity *can* be objectified it can never be entirely so. Having briefly considered points in favour of the first (less contested) claim, the latter two must now be defended and their relevance to the evaluation of higher-order theories made clear.

Higher-order theories claim that inner awareness is a matter of reflection or introspection. The problem for this view is that it seems wrong to say that we are only ever conscious of our experiences when reflecting upon them. Zahavi succinctly expresses this subsection of the objectification argument in arguing that "[my] pre-reflective access to my own mental life in first-personal experience is immediate, non-observational and non-objectifying" (Zahavi, 2006, p.6). What he means when he claims that it is "non-objectifying" is that he "[does] not occupy the position or perspective of a spectator or in(tro)spector on it" (Zahavi, 2006, p.6). The sense of being a spectator can come about owing to our ability to direct our attention towards various aspects of our experiences, as we have seen. This is a misinterpretation of the nature of introspective experience. To further claim that a subject objectifies *all* of its occurrent experience in such a manner is to either accept that unattended conscious states are non-experiential (which is to (implausibly) deny the existence of the phenomenal background) or it is to hold that a subject occupies the position of spectator at every moment of their experiential life. This latter entailment is at least as problematic as denying the phenomenal background: it utterly mischaracterises vast portions of our streams of consciousness. Zahavi describes this kind of reflective consciousness as "a detached objectifying self-awareness that (normally) introduces a phenomenological distinction between the observer and the observed" (Zahavi, 2005b, p.21).

If you pause for a moment and once again reflect upon your current conscious experience as a whole, with a little practice and some effort you might be able to seemingly adopt the perspective of a mere observer of it. Although difficult, it is possible to get the sense that your experience, considered as a whole, is something entirely distinct from you, which you are able to passively observe. Even if sense can be made of such an impression it is not at all what it is *typically* like to enjoy experiences. Our attention is usually on the objects, people and states of affairs that we are dealing with or interested in. The experience itself is something we simply consciously *live through*, not something we adopt the perspective of spectator on unless we are purposefully introspecting in such a manner. This is perhaps most

obvious when we consider less specific elements of our experience, such as our overall mood. We can sometimes be only dimly aware that we are in a particular mood, even though its effect on the overall character of our stream of consciousness can be pervasive. We can be in the midst of it, utterly in its grip, yet not be explicitly reflecting on it. To claim that we are observing this and all other aspects of our experience as an objectifying spectator is to misdescribe the situation entirely.

As Zahavi notes, "in my everyday life, I am absorbed by and preoccupied with projects and objects in the world, I am not aware of my own stream of consciousness as a succession of immanent objects" (Zahavi, 2011, p.17). This, as it stands, is a purely phenomenological claim to which, arguably, there is only one *direct* way of evaluating its veracity: by attending to one's own experience often and diligently in an attempt to ascertain which model of description is most accurate. Claims such as this are abundant in the philosophy of mind and this basic call to "look and see" has been expressed in a number of different ways. This is Reid's version:

> In order, however, to our having a distinct notion of any of the operations of our own minds, it is not enough that we be conscious of them; for all men have this consciousness. It is further necessary that we attend to them while they are exerted, and reflect upon them with care, while they are recent and fresh in our memory. It is necessary that, by employing ourselves frequently in this way, we get the habit of this attention and reflection; and, therefore, for the proof of facts which I shall have occasion to mention upon this subject, I can only appeal to the reader's own thoughts, whether such facts are not agreeable to what he is conscious of in his own mind. (Reid, 1854, p.57)

There are many, no doubt, who would regard this kind of evidence as suspect at best and patently unreliable at worst. It seems, however, that if we are to take consciousness seriously as something to be investigated and explained then this kind of evidence (personal confirmation) is somewhat unavoidable. This is a result of taking a "consciousness-first approach" (Goff, MS) as opposed to a "brain-first approach". This means accepting that the physical sciences are needed in order to explain consciousness, but also holding that "our first person grasp of consciousness should shape, rather than be shaped by, our scientific picture of matter" (Goff, MS). A theory of consciousness can have (and, indeed, almost certainly will have)

counterintuitive implications, but it must also respect personal confirmation at a fundamental level. Suppose a grand theory of consciousness was put forward based on solid third-person evidence. Further suppose that one of its implications was that seeing green and seeing red are qualitatively indistinguishable experiences for human subjects (or that all people with your colour of hair are philosophical zombies). In this (unlikely) scenario it would be most reasonable to reject such a theory based on nothing but personal confirmation from the first-person perspective. An honest attempt to explain consciousness will, in the end, need to match up with many central aspects of our subjectively attained knowledge of conscious experiencing (otherwise, it will be explaining something else).[15] However, even if such an appeal to personal confirmation is not wholly without merit it would certainly be better if we had more reliable means of evaluation with which to back up the original phenomenological claim. As it turns out, this subsection of the objectification argument (the claim that we are not typically aware of our experiences by way of reflection) finds strong support in the regress argument, as shall be detailed shortly.

Firstly though, the final strand of the objectification argument needs consideration: this is the claim that a moment of experience can never be *entirely* objectified, even if some element of it can be. A subsection of the objectification argument is relevant here (which itself lends support to all three of the objectification argument's principle claims) and can be called "the first-person argument" (MacKenzie, 2007). Consider your current experience. Intellectual honesty demands acceptance of the following claim: you might be very mistaken in your beliefs about the world. You might, in fact, be (almost) entirely mistaken. It could be that your experience is being caused not by the world you take yourself to be in but by a Cartesian demon, an advanced computer simulation or some other unknown power. That we cannot know either way need not be taken to be a "scandal to philosophy" (Kant, 1934, xi). Instead, we can just accept this as a basic fact about our epistemological situation (perhaps a universal fact for all conscious beings, necessarily entailed by the fundamental division between an experiencing subject and the world its experience presents). That this is our situation is indeed recognised by most. Granting, then, that your current experience might be utterly misleading with regards to the nature of the external world, notice that there is

[15] This claim is not, of course, committed to the mistaken view that *all* of our introspective knowledge is accurate.

at least one element of your experience that you cannot be wrong about: that it is *yours* (setting aside, for now, the question of precisely *what* you are) and not (exclusively)[16] someone else's. You might be misidentifying the objects in front of you, perhaps even all of them, but you cannot misidentify yourself as the bearer of the experience. Descartes was right: hyperbolic doubt does not stretch this far. Shoemaker puts the point as follows:

> The statement "I feel pain" is not subject to error through misidentification relative to 'I': it cannot happen that I am mistaken in saying "I feel pain" because, although I do know of someone that feels pain, I am mistaken in thinking that person to be myself. But this is also true of first person statements that are clearly not incorrigible; I can be mistaken in saying "I see a canary," since I can be mistaken in thinking that what I see is a canary or (in the case of hallucination) that there is anything at all that I see, but it cannot happen that I am mistaken in saying this because I have misidentified as myself the person I know to see a canary. (Shoemaker, 2003, p.8)

The rest of the first person argument follows quickly on: "identification goes with the possibility of misidentification" (Shoemaker and Swinburne, 1984, p.103) and as it is not possible to misidentify yourself as the bearer of an experience, your awareness of yourself as the bearer of your experience cannot be the result of identification or recognition. This is a problem if you hold that you are aware of yourself as the subject of your experience by way of some perception-like awareness of an inner object.[17] The first person argument shows that even if *an* aspect of your experience is objectified (where a *seeming* "phenomenological distinction between the observer and

[16] It *might* be that you are wrong in believing yourself to be the only subject of your current experience: one of the more intriguing interpretations of split-brain phenomena explores the possibility that even normal human subjects have a second subject of experience residing in their brain (one who is unable to communicate this fact). We cannot know, from our own pre-theoretical perspective, that this is not the case. Such a subject *might* also be aware of some of your experiences (depending on whether or not experiences are the kind of things that can be shared in such a way). But this curious possibility does not change the basic point: you cannot be mistaken that your current experience is indeed *yours*, and not solely someone else's.

[17] Consider, by analogy, the act of perceiving oneself in a mirror for the first time. There is nothing about the mirror image itself that can identify it as an image of you unless you are already aware that it is *you* who perceives it. The same problem arises when the "mirror image" is held to be an *objectified* aspect of an experience identifying its bearer.

the observed" (Zahavi, 2005b, p.21) occurs) *not all of it is*; namely your awareness of yourself as the bearer of it (or, less contentiously, your awareness of its occurrence in *this* particular stream of consciousness, as opposed to some other stream)[18]. Furthermore, it shows that objectification is not the *typical* structure of inner awareness (for you are *always* aware that your experiences are yours in the sense that they are occurring *here;* in *this* stream of consciousness). The upshot of these arguments is that at least *this* aspect of our inner awareness is immediate and *pre*-reflective. It is not the result of a secondary act of inner perception, directed at a distinct inner object. As we will see in §2.8 there are also more general, and fundamental, reasons to deny that inner awareness is perception-like.

§2.6 The reportability argument

Right now your attention is, hopefully, focused on the words you are reading and the meaning they seek to convey. You are not (at least, until you read this sentence) focusing on yourself or on what it is like for you to be reading this. Yet if someone were to interrupt your concentration and ask you what you were doing, you would be able to immediately report that you were engaged in the activity of reading. This ability to instantly report on your experience extends to every kind of conscious activity that you might be engaged in (aside from, of course, when you are asleep and unable to respond to any environmental stimuli, or in some other way incapacitated). This is the plausible starting claim of the reportability argument. Further to this, your report would not be the result of, firstly, hearing the question and then studying your current situation in order to determine what you had been doing. You would not scan your environment and, upon seeing a document in front of you, *infer* that you had been reading. Your knowledge of your current activity was already available to you, prior to any investigation of your context, and thus allowed for an immediate report. The reportability argument holds that the explanation for this is that you were already conscious of your experience prior to the question, which allowed you to instantly recall it and report it. Given that you were not explicitly attending to the experience (you were focused on understanding the text) it follows that introspective attention (or

[18] We will return to this consideration in §6.3.

reflection) is not necessary for inner awareness. You had a *pre*-reflective consciousness of your experience. Therefore a distinct higher-order awareness, directed at another state in an act of reflection, is not required for inner awareness to occur. This is the essence of the reportability argument. Zahavi elaborates:

> the self-consciousness on the basis of which I answer the question is not something acquired at just that moment, but a consciousness of myself that has been present to me all along. To put it differently, it is because I am pre-reflectively conscious of my experiences that I am usually able to report immediately, that is, without inference or observation, if somebody asks me what I have been doing, or thinking, or seeing, or feeling immediately prior to the question. (Zahavi, 2005b, p.21)

The latter sentence is better for our current purpose: it does not make use of the easily misunderstood phrase "self-consciousness" and does not presume that the inner awareness we are currently interested in is synonymous with, or always entails, a consciousness of oneself (this requires further argument, which is forthcoming in §3). As should be obvious this Sartrean position on the issue of inner awareness is in radical disagreement with higher-order theories. Armstrong, for example, holds that inner awareness is the result of our introspecting our mental activity. For Sartre, it is in fact exactly the other way around: our pre-reflective inner awareness is what *allows* for reflection in the first place: "it is the non-reflective consciousness which renders the reflection possible" (Sartre, 2010, p.9). We can reflect in such an immediate way precisely because we were already conscious of our experiences, albeit often non-focally. In other words, our awareness of any given experience is always already there: it is not the result of an act of reflection. This is not, however, an argument against a perceptual-like higher-order theory such as Armstrong's (although, as we have seen, his theory faces other serious problems). He too can account for the data of immediate reportability: he can claim that our ability to report on our experiences without inference is the result of an accompanying inner perception of them. What the reportability argument does do however, in conjunction with the following regress argument, is force the hand of another kind of higher-order theory into making radical, and questionable, claims concerning the nature of inner awareness.

§2.7 The regress argument

The objectification argument's various attacks on a perceptual or perceptual-like model of inner awareness have been influential. In an effort to avoid losing the theoretical benefits of a higher-order theory of consciousness and inner awareness (namely; its compatibility with a reductive account of consciousness) a different take on higher-order awareness has been suggested (one that does not posit the problematic notion of inner *perception*). In this version the higher-order awareness in question is a thought or a mental representation of some kind. Rosenthal is arguably the most prominent exponent of this kind of view. His starting assumption is strong:

> There is a natural way of understanding how conscious states differ from mental states that are not conscious. No mental state is conscious if the individual that is in that state is in no way aware of it. (Rosenthal, 2012, p.2)

This much seems easy to agree with. The intrinsic-inner awareness view can (and should) concur with this starting point. What is, however, very much open to dispute are the following claims from Rosenthal:

> We are conscious of our conscious mental states by virtue of having accompanying thoughts about those states. When a mental state is conscious, we are transitively conscious that we are in that state. So the HOT [higher order thought] that accompanies it will be a thought to the effect that one is in the target mental state. Because these thoughts are about other mental states, it will be convenient to call them higher order thoughts. (Rosenthal, 1994, p.361)

Rosenthal identifies three distinct kinds of consciousness: creature consciousness, transitive consciousness and state consciousness. Creature consciousness is "roughly, the opposite of being asleep or knocked out" (Rosenthal, 1994, p.355) and transitive consciousness occurs when such a creature perceives or thinks of an object. Rosenthal's reason for calling this "transitive" consciousness is as follows:

> A full description of a creature's being conscious of something always involves

> mentioning the thing the creature is conscious *of*. So it is natural to call this property *transitive* consciousness. (Rosenthal, 1994, p.355)

For Rosenthal, being transitively conscious of *x* means being in a mental state that represents *x* (Rosenthal, 1994, p.356). And, finally, state consciousness is that property which distinguishes conscious mental states from unconscious mental states. Although Rosenthal does not use such language it seems natural to hold the concept of state consciousness, as defined, to be referring to the property of phenomenality. There is nothing it is like to be in an unconscious state precisely because it has no phenomenology. So on the higher-order thought view when a mental state is phenomenally conscious this is because its subject currently has a thought about that state. The subject is transitively conscious of the target state, in virtue of which the mental state in question is phenomenally conscious. Before considering this theory specifically with reference to inner awareness, it will be helpful to see how the regress argument tackles its more general goal of accounting for the difference between conscious and unconscious mental states.

The regress argument seeks to force a higher-order account of Rosenthal's kind to choose between two equally problematic horns of a dilemma (whilst also claiming that an intrinsic-inner awareness view faces no such difficulty). The argument can be condensed as follows. On Rosenthal's theory what makes a mental state conscious is its being the target of a distinct higher-order thought. The following question can then be posed: is the HOT conscious or not? If it is conscious then, in order to be consistent, we must posit yet a further state directed at the HOT (an even higher HOT) to account for *its* being conscious (given that being the target of a higher-order thought is held, on this view, to be what a mental state's being phenomenally conscious consists in). At which point, the same question can then be asked of the second HOT, and so on. We now face an infinite regress that would require an infinite chain of higher-order thoughts to accompany each and every one of our world-directed conscious states.

Not only is this picture completely at odds with how the general character of experience strikes us in having it (there is not the slightest hint of an infinite hierarchy of states in our phenomenology), but also such a view would clearly not sit well within any reasonably naturalistic perspective on the relationship between the brain and consciousness. It is justifiable to hold that, although we may be ignorant of the

precise relationship between the brain and consciousness, we can be fairly confident that the brain is intimately involved in the production of the experiences we enjoy. To hold that the lump of matter inside our skulls is capable of producing an infinite chain of mental states for *every* conscious one is absurd. Thus, HOT theorists cannot, and do not, hold that the original higher-order thought is itself conscious.[19] Instead, they must hold that it is unconscious and, as we shall see, this claim has serious problems of its own. This, in a nutshell, is the regress argument. That there is a choice that the HOT theorist must make seems evident.[20] Whether or not it is a *dilemma* depends entirely on the plausibility of holding higher-order thoughts (or representations) to be unconscious. This issue now needs addressing.

An immediate problem facing a higher-order thought theory of *this* kind concerns the issue of animal and infant consciousness. The HOT theory claims that, in order for a state to be phenomenally conscious, an unconscious thought about that state is required. However, even a relatively simple thought along the lines of "I am seeing a tree" demands a certain level of mental sophistication and linguistic mastery that is unlikely to be present in non-human animals and human infants. The question of animal and infant consciousness is, of course, a difficult one: we know of other people's conscious experiences in large part because they tell us about them, and this option is unavailable to us when we are concerned with conscious beings who cannot speak. Having said as much, it seems quite reasonable to suppose that at least *some* non-human organisms (and human infants) are capable of conscious experience. There are plenty of persuasive instances that we can bring to mind: the shame a pet dog seems to suffer when caught in the middle of a forbidden act is a particularly powerful one, and difficult to make sense of if we are restricted to admitting the existence of "shame-behaviour" only. This intuition is especially powerful when we consider species whose neuronal architecture is closer to our own, in terms of its evolutionary history and functional structure (or, indeed, younger members of our

[19] Even if we were to suppose that, somehow, the HOT is conscious without requiring a further conscious HOT there is another problem: it does not seem reasonable to hold that we are consciously thinking about *every* conscious experience we undergo - this would entail that "consciously thinking about something other than one's current experience (e.g. pondering some abstract subject matter) would plunge one into darkness and silence" (Dainton, 2004a, p.12) and this is clearly not the case.

[20] There is, of course, a third possible choice: hold that inner awareness is intrinsic to the ("target") conscious mental state, and thus cut the regress off before its very first step. This choice is unavailable, however, to someone committed to a higher-order view.

own species).

The HOT theorist can bite the bullet and deny this powerful intuition, but a more convincing move is to modify the theory: it is not a higher-order *thought* that is responsible for conscious experience, but an intentional state or mental representation of some suitable kind. So long as the supposed representational state does not require conceptual capabilities beyond that of all non-human animals and human infants, a higher-order representation (HOR) theory can avoid such counterintuitive claims regarding their conscious lives (or lack thereof). But, as we shall see, this modified version of the higher-order model is itself fundamentally problematic.

Firstly, there are yet more problems with holding the higher-order state in question to be a thought. These concern the issue of misrepresentation (and, as such, can also speak to the non-linguistic intentional state versions of the HO model, if those states are held to represent their target). The first issue concerns inference: I may be able to infer that I am in a certain conscious state by observing my behaviour or by being told by someone else. I would, as a result, have a thought about that specific state. However, it would clearly not be in virtue of this inference-based thought that its target state is conscious. A HO theory, without amendment, would wrongly claim otherwise. Not surprisingly, an amendment is now usually offered to the effect that the relevant thought is not arrived at inferentially. It seems, however, that this is not much more than a stopgap: in attempting to explain phenomenal consciousness in terms of the relation between a distinct representation and its target state, the possibility of radical misrepresentation is an unavoidable threat. To be sure, there are ways we can be mistaken regarding our experiences. But, as Block shows, the *kind* of misrepresentation possible on HO views is problematic:

> Suppose that at time *t*, I have an assertoric higher order thought to the effect that I am experiencing seeing something green, but in fact I am having no visual representation at *t*: the thought is 'empty'. Let us suppose further that the higher order thought is not arrived at inferentially. Also, I have no other higher order thoughts at *t*. The theory supplies a necessary and sufficient condition for a conscious episode. An episode is conscious at *t* if and only if it is the object of an assertoric higher order thought at *t*, arrived at non-inferentially. The sufficient condition dictates that this thought at *t* is sufficient for a conscious episode at *t*. By the necessary condition, that conscious episode at *t* is the object

> of a simultaneous higher order thought. In the example, there is only one higher order thought at *t*, and we can assume it is not self-referential. So there is no conscious episode at *t* after all. *Thus, the sufficient condition and the necessary condition are incompatible in a situation in which there is only one non-self-referential higher order representation*. (Block, 2011, pp.424-5, original italics)

In others words, the theory is incoherent. If accurate, this is a powerful objection and there are only a limited number of ways around it. The first would be to hold that the conscious state in question is in fact the higher-order thought *itself*, or that the higher-order thought is itself somehow *part of* the conscious state. This, however, would no longer be a *higher-order* thought at all. Alternatively, one might claim that although there is no visual representation there is still a conscious experience "instantiated only in intentionally inexistent states" as Block puts it (2011, p.425). At this point, it seems that we have departed from taking consciousness seriously.

Block's treatment of the issue can, however, be called into question:

> But Rosenthal's theory is *not* vulnerable to this objection, as Rosenthal doesn't endorse the stated necessary condition. He is happy that lone, targetless HOTs supply the subjective mental appearances characteristic of consciousness. Block's necessary condition is really only necessary for a *pre-existing sensory* state: its only hope of entering consciousness is via HOT representation. But HOTs are ultimately responsible for generating subjective mental appearances. So Block's attempt to disintegrate HOT theory fails. (Coleman, 2015, p.4)

As Coleman goes on to say, however, "there's surely *something* solid within the residue of dissatisfaction around Rosenthal's treatment of HOT mistargeting" (Coleman, 2015, p.4). The issue of misrepresentation also comes into view when we consider what HO theories have to say about the nature of inner awareness. For Rosenthal, when the (typically unconscious) HOT, in virtue of which we have a conscious experience, is *itself* the target of a further higher-order thought, inner awareness is present. The reportability argument puts pressure on this view: given that we are typically able to instantly report on our conscious activity Rosenthal would have to hold that we are almost *constantly* thinking about our experiences whenever we are having them. The phenomenal background, in its diversity and depth, stands

strongly against such a claim.

Further to this, applied as a higher-order take on *inner awareness*, this view faces the same fundamental problems that the general higher-order take on conscious experience suffers. It also has a new one:

> If one is introspecting, that is, directing one's attention to one's own mental states, then, according to Rosenthal, one has a (so far) unconscious third-order thought whose intentional object is the second-order thought with the content that one is in S. The content of the third-order thought, however, is not that one is in S, but that one thinks that one is in S. If S is a perceptual state of seeing the blue sky, then the content of the third-order thought would be that I think I am seeing the blue sky. This is the content that should be conscious. But, in fact, what is really conscious when we are introspecting are contents of the form 'I am now in S'. We may be focussing on our focussing on mental states too, and then we would have conscious contents of the form 'I think I am in S'. But in all cases of consciousness of first-order mental states like perceiving, feeling or desiring, the form of content is 'I am now in S'. Thus, it seems that explaining introspective consciousness by third-order thoughts leads to a wrong prediction as to the content which is conscious. (Schroder, 2001, p.38)

It is tempting to suppose that there is a fundamental problem with the higher-order model. In all of its perceptual, thought-based and representational forms considered here, and in its tackling of both phenomenal consciousness in general and inner awareness more specifically, it seems forced into making one problematic claim or another. There are, of course, many kinds of higher order theories in the literature - each with their own strengths and weaknesses. Some, such as Coleman's (2015) "Quotational" higher-order thought theory, deal with the problem of mistargeting, for example, in a more optimal fashion than Rosenthal's. But one might be forgiven for thinking that there is something fundamentally faulty with the model as a whole. Dainton expresses this general concern well:

> even if objections such as these do not prove insuperable, there is a difficulty, or perhaps more accurately, a deficiency, of a more fundamental kind. The prime goal of the HOT theorist is to provide a coherent and believable account of the

> difference between conscious and non-conscious mental states. Evidently, this is a non-trivial task, as there is a world of difference between the two (just think what it is like to regain consciousness). The HOT theorist would have us believe that all hangs on the presence or absence of non-conscious intentional states. I fail to see how this can be the complete story. (Dainton, 2004a, p.14)

Considering the wealth of objections that have been put to the various higher-order views, it would be unreasonable to blame such skepticism on prior ontological commitments or a lack of imagination (on the contrary, it seems that one of the few powerful motivations for defending a higher-order view is its compatibility with a reductive account of conscious experience: a common background commitment). If we honestly accept the reality of conscious experience on its own terms it can seem likely that higher-order views are on the wrong track:

> Suppose HOT theorists are correct, and all manner of mental states can exist in both conscious and non-conscious modes. Consider two subjects, S1 and S2, both of whom are having experiences (or sensory states) caused by looking at a red balloon; call these E1 and E2. Whereas E1 is conscious, E2 is non-conscious. In virtue of this difference, the overall consciousness of S1 will be different from that of S2: S1 will be aware of an intrinsic quality, phenomenal redness, S2 won't be. The HOT theorist explains this intrinsic difference in relational terms: S1 is conscious of phenomenal red by virtue of possessing appropriate unconscious intentional states, intentional states that S2 lacks. But the intrinsic difference between what it is like to be S1 and what it is like to be S2 surely cannot *consist* in a relational difference, or at least, not in *this* relational difference (i.e., being differently related to *non*-conscious states). (Dainton, 2004a, p.14, original italics)

There are, then, a number of problems that higher-order theories face in their attempts to account for phenomenal consciousness generally and inner awareness specifically. In particular, the effort to explain intrinsic differences in merely relational terms appears fundamentally problematic. It goes without saying that the current discussion is not a fully exhaustive treatment of the issue: such a project would require far more space than is available here, given the other arguments and claims I wish to make.

Having said as much we have seen a host of problems that can motivate a desire to look elsewhere for convincing answers. In this respect, the *way* in which higher-order theories get into trouble might help us to form a clearer picture of the issues at hand. At bottom, their problems arise from positing a *distinction* between a conscious state and the subject's awareness of that conscious state.

§2.8 An alternative to the higher-order model

This basic idea is at work in each of the perceptual, thought-based and representational versions and is used to explain the difference between conscious and unconscious mental states, as well as to account for inner awareness. We may be in a better position if we reject this underlying claim. With reference to consciousness considered generally, the lesson to be learnt is as follows:

> consciousness is inseparable from phenomenal contents: when a given phenomenal item comes into being, it comes into being as a conscious experience; to be an experience it does not need to fall under any separate awareness … In other words, contents are themselves intrinsically conscious, and hence - in a manner of speaking - they are self-revealing or self-intimating. That is, phenomenal contents become conscious simply by coming into existence. Whenever phenomenal properties are realized, or phenomenal objects come into existence, conscious experience occurs. I shall call this non-dualistic model of consciousness the *Simple Conception* of experience. (Dainton, 2006, p.57)

This model avoids the various pitfalls of the higher-order approach. We can condense the claim as follows: experience is self-intimating. There is, however, one problem that it faces which a higher-order view seemingly avoids: it is difficult to see how we can reductively explain conscious experience if it is held to be self-intimating.[21] To claim, however, that only reduction-friendly theories are to be put on the table is to

[21] There is good reason to think that the self-intimation of experience is in fact the most fundamental aspect of the mind-matter problem: the *true* source of the explanatory gap.

beg the question in a serious and unhelpful way. The *Simple Conception* avoids the many problems facing the higher-order theories, and is perfectly in keeping with our basic phenomenological data. If it makes the reductive explanatory project seem more difficult then this is a problem for the reductive approach, not the *Simple Conception*. The hard problem of consciousness exists for everyone (excluding eliminativists and illusionists, who have their own difficulty to deal with)[22]; the *Simple Conception* merely identifies and admits of an important aspect of it.

The *Simple Conception* avoids the problems leveled at higher-order theories because of its refusal to posit a separation between phenomenal content and awareness. Dainton presents his account in stark opposition to the notion of awareness detailed here:

> Awareness cannot itself be observed, it is not an object, not a thing. Indeed, it is featureless, lacking form, texture, colour, spatial dimensions. These characteristics indicate that awareness is of a different nature than the contents of the mind; it goes beyond sensation, emotions, ideation, memory. Awareness is at a different level, it is prior to contents, more fundamental. Awareness has no intrinsic content, no form, no surface characteristics - it is unlike everything else we experience, unlike objects, sensations, emotions, thoughts or memories.
>
> Thus experience is dualistic, not the dualism of mind and matter, but the dualism of awareness and the contents of awareness. To put it another way, experience consists of the observer and the observed. Our sensations, our images, our thoughts - the mental activity by which we engage and define the world - are all part of the observed. In contrast, the observer - the 'I' - is prior to everything else; without it there is no experience of existence. If awareness did not exist in its own right there would be no 'I'… no transparent centre of my being. (Deikman, 1996, p.351)

I will follow Dainton in labeling this concept *awareness**. Its fundamental characteristic is its distinction from whatever content it happens to be observing. The above description can in some ways seem convincing: that there is an aspect of conscious experience that is not easily described as a "thing" or an "object" is not an

[22] Namely: the self-evident existence of conscious experience.

entirely objectionable claim. Such thinking might go some way to explaining the historical popularity of the notion of an immaterial soul. It can also become even more convincing when one engages in certain meditative practices: “neti neti” (not this, not that) is one such practice and a rough instruction can be put along the following lines.

Consider your current experience. Notice that it seems obvious to you that *you* cannot be any of the worldly objects that you are currently conscious of: they are distinct from you and this is evident to you on the basis of the fact that you are *here* observing them, and they are *there* being observed. This seems to hold regardless of the veridical status of your perceptions: even if the objects are in fact hallucinated objects, or part of a virtual world, they are seemingly distinct from you as the observer of them. Now consider your body: the same basic fact seems to apply. Although there are important differences in the way your body figures in your phenomenal sphere - for one thing it is constantly present in a way that other objects are not - it is still something you can observe, and therefore something experientially distinct from yourself as the observer. Now consider your entire inner mental life: your thoughts, sensations, emotions and, even, the sense of being the person that you are. All of this, so the argument goes, is something you can observe. Such observation is different from the perceptual kind, but it is still an act that consists of an observer and something observed. In short: if you are the one who is observing *x* then you cannot *be x* owing to the fact that *x* has the property of being observed, whereas *you* have the property of observing.

Having followed such instructions the following insight is purportedly available: you, as the observer, can *never* be observed. If you think you have managed to observe yourself as awareness, you have in fact constructed a subtle image of yourself: *you* are always one step behind the observed phenomena. We have arrived at awareness*. Such a model can indeed be attractive: it is possible, although difficult, to enact the above practice and be met with the impression that, as an observer, one is something akin to an invisible witness. As will be seen in sections §6.4 and §6.5, however, there are good reasons to think that such an impression can be explained, and even utilised, without positing a distinction between awareness and content. Such a reappropriation is necessary for the following reason: it is not at all obvious that such an alleged distinction between content and awareness is in fact coherent. This will become clear by examining the four possible options available to an awareness*

defender, identified by Dainton as follows:

S1 awareness* cannot exist independently of content, and content cannot exist independently of awareness*

S2 awareness* cannot exist independently of content, but content can exist independently of awareness*

S3 awareness* can exist independently of content, and content can exist independently of awareness*

S4 awareness* can exist independently of content, but content cannot exist independently of awareness* (Dainton, 2006, p.48).

Consider S1: on this view content cannot exist independently of awareness. Why should this be so? Dainton asks us to consider some phenomenal content: a token of blueness. Picture a simple blue region of a visual field. What reason do we have to think that this can only exist in the presence of an awareness*? Remember that awareness* does not, in itself, possess any phenomenal characteristics: as Deikman argues it is "featureless" and "has no intrinsic content" (Deikman, 1996, p.351). What then does it bring to the blueness in order to make it exist? There is no blueness or "colorfulness" to be found in awareness*. This problem is exacerbated when we consider the sheer *diversity* of the phenomenal content that we enjoy:

> If awareness* is in itself diaphanous, perfectly transparent, bringing the same intangible illumination to all its objects, how can it bring such diverse phenomenal properties as colour and sound into the world? Supposing a transparent awareness* could be responsible for phenomenal diversity is as absurd as thinking one could convert a television from black-and-white to colour by holding a sheet of plain glass in front of the screen. (Dainton, 2006, p.49)

Think of the dramatic phenomenal difference that exists between a patch of blueness and a loud ringing noise, or between a feeling of nausea and the smell of wine. How

could a featureless awareness be responsible for such different kinds of content when it is itself devoid of all of these kinds of properties? Wolfgang Fasching (2009) finds the above consideration unconvincing, and argues that "[one] could equally ask what ingredient extension brings to colour and conclude from the fact that it does not contribute any colour-shading to it that colour can do just as well without extension" (Fasching, 2009, p.144). This, however, does not seem entirely analogous: Dainton is not claiming that content can do "just as well" without awareness *full stop*, if this is taken to mean that phenomenal content can exist in the absence of consciousness. He is claiming that phenomenal content can do just as well without a distinct and *featureless* awareness. The *Simple Conception* of experience does not claim that phenomenal content can exist unconsciously: it is not without consciousness or awareness in *this* sense. Dainton's claim is that the concept of awareness* does not successfully refer to anything distinct from the intrinsic nature of phenomenal contents themselves. In terms of Fasching's analogy, a "*Simple Conception of Extension*" would assert the following: extended contents, including colour, can do just as well without a *featureless* (and thus unextended) "space" (or space*). This would be a reasonable position to hold. S1, then, seems dubious. The next option needs examining:

S2 awareness* cannot exist independently of content, but content can exist independently of awareness*

On this view, content does not require awareness* in order to exist. The role of awareness*, then, is not to create content but to *reveal* it. As Dainton argues, this view is just as problematic as S1, if not more so, in that it renders awareness* fundamentally *superfluous* (Dainton, 2006, p.50). Consider all of the phenomenal content that is currently determining what it is like for you to undergo the particular experience you are enjoying at this moment. If this content can exist independently of awareness* (as is claimed in S2) then what difference would it make should this awareness* suddenly cease to exist (Dainton, 2006, p.50)? The phenomenological character of your current experience would remain *unchanged*. Dainton goes on:

> even if a pure awareness* is now gazing down upon your consciousness and is in some manner apprehending the character of your experience as it unfolds,

> what has this entity got to do with you? The answer seems to be: nothing whatsoever. The idea that you would cease to experience anything if this awareness* were to vanish seems quite absurd. Once again the difficulty lies with the featureless character of awareness*: if awareness is wholly without phenomenal features it is hard to see that anything would be lost if it were to disappear, and it is very hard to see why experience itself should be impossible in its absence. (Dainton, 2006, p.50)

It seems, then, that both S1 and S2 are problematic. What of S3 and S4? Both of these views hold that awareness can exist without content. A fundamental problem faces this view and is neatly presented by Dainton when he asks "[what] would a wholly contentless awareness* be like? … What would differentiate a bare awareness from nothing at all?" (Dainton, 2006, p.53). What it is like to experience *x* is in part determined by the phenomenal content involved. If *x* is an awareness that is itself empty of all phenomenal content, then so too is the "experience" of *x*: a contentless awareness would leave no impression at all. In short, a truly featureless awareness is not something that we could become aware of. We can put the issue in terms of a rhetorical question: what would this pure observer (or witness) be like in itself? As we will investigate in §6 certain meditative techniques espoused in Indian philosophy are geared towards facilitating a self-recognition on the part of their aspirant. The practitioner is guided into recognising their essential nature as consciousness. Consider this: what would be the point in achieving such a realisation if one's nature was not *like* anything at all? Crucially: how would you even know when you *had* achieved the desired recognition if such an awareness* is incapable of showing up in your experience? It is worth keeping in mind the distinction between objectless and contentless experience:

> It has been reported that in very advanced states of meditation … a mode of pure *objectless* conscious experience can be attained where there is *something it is like* to be in such a state. If such modes are really possible, then they would suggest that this subject-awareness has its own intrinsic phenomenal character - always present but largely unnoticed in ordinary conscious states. (Albahari, 2009, p.64, original italics)

Clearly such an awareness is not *contentless*: there is something it is like to reach such a state of consciousness. Indeed, it is presumably quite an enjoyable state of mind given the efforts many undergo in an attempt to reach it. Emptying one's mind of all object-directed thoughts and perceptions would indeed be an *atypical* experience: it might strike us as considerably less cluttered, and thus further towards "contentless", than our usual conscious states, but in order for it to be worth pursuing it must have *some* qualitative character.

It seems, then, that separating content from awareness causes serious problems for both the HOR approach and the featureless awareness account. Instead of positing such a distinction we should agree with Dainton that experience is *self-intimating*: it is not revealed by a distinct awareness but instead reveals itself. It is crucial to note that self-intimation is a different, but related, concept to "self-consciousness", even when "self-consciousness" is taken to refer only to inner awareness (i.e. not to a conceptualisation of oneself). A defender of self-intimation might wish to claim that experience is self-intimating but that it lacks inner awareness. As has been shown, however, this latter impulse needs to be rejected.

We have covered a lot of shifting ground, and it will be useful to offer a general analysis of the situation. The view that inner awareness is intrinsic to experience, as we saw, can be motivated by an appeal to the pre-reflective nature of consciousness, in opposition to reflective and higher-order views. Many have claimed that such considerations speak directly to the issue of what we *are*. Sartre (2010), for one, claims that the pre-reflectivity of awareness entails that our nature is "being-for-itself" (consciousness). I think such an impulse is right, and that - in fact - such an entailment can be defended merely on the basis of accepting two plausible views: Dainton's *Simple Conception* of experience and UIA. Such a simple presentation of the issue might make for a more persuasive case for those who are suspicious of, or find it understandably difficult to follow, Continental Phenomenology's terminological style. More importantly, its simplicity might shed light on just why inner awareness has such implications concerning our essential nature.

To take stock then, and allow ourselves access to only the minimal amount of concepts necessary for the forthcoming argument to hold: conscious experience is self-intimating and inner awareness is a ubiquitous feature of conscious experience. It therefore follows that, if we momentarily take "self-consciousness" to mean *nothing more than* "inner awareness", we can concisely summarise our findings into the

following claim: **conscious experience is necessarily both self-intimating *and* self-conscious**. This is just to say that, for any given experience (*e*) of a subject (*s*); *e* is self-intimating (it doe not require a distinct awareness or HOR in order to be conscious) and *s* is aware of *e* (i.e. inner awareness is present). It is important to note that this formulation makes no claims about the subject's awareness of *itself*. However, as I have hinted, these simple premises have an interesting implication concerning just that. This now needs addressing.

3: Being Consciousness

Having done the necessary work we are now in a position to advance a core argument. Simply put, if we accept the claims previously defended then we must also accept the following (preliminary) thesis:

The Identity Thesis: the subject of an experience (*e*) is identical with *e*.

The Identity Thesis is, of course, not particularly new. What *is* of interest is that only a minimal amount of (plausible) premises are needed for it to be entailed. It does not depend on the notion of "mineness" (or "for-me-ness"), although it need not deny the existence of such a feature of consciousness (depending on how "mineness" is interpreted). Given that the notion of "mineness" is open to serious attacks (an issue we will cover in §6.3), avoiding the need for it affords the argument a good deal of strength. We can call it "The Exclusivity Argument" and its most succinct form is the following:

P1. Any given experience (*e*) is self-conscious and self-intimating
Therefore:
C1. *Only e* is conscious of *e*
Therefore:
C2. The subject of *e* is *e*

The first premise has already been argued for and found to be a reasonable claim, so long as the term "self-conscious" is taken to mean no more than that inner awareness is present, i.e. that the subject of *e* is conscious of *e*. C2 follows from the conjunction of C1 and the first clause of P1 quite obviously, given this definition of self-consciousness. The crucial element left to defend, then, is the entailment of C1 from P1.

This is most easily done by starting with the first clause of P1: any given experience (*e*) is self-conscious. Recall that "self-consciousness" so used is not referring to any *concept* of self that may or may not be present in *e*. Neither is it

referring to an awareness of the property of "mineness". According to the foregoing discussion, *e* is self-conscious just when the subject of *e* is conscious of *e* (i.e. inner awareness is present). The key claim, then, is that C1 is entailed when this is accepted along with P1's second clause. Recall the *Simple Conception*'s view that experience is self-intimating. Once again the "self" in the phrase "self-intimating" should not be taken to imply, on its own, an awareness of "a self" or an awareness of "itself". To say that experience is self-intimating is, in one sense, a merely negative claim: it is to claim that phenomenal properties do not require a distinct awareness in order to become manifested. Experience is, metaphorically speaking, "self-illuminating". Perhaps less contentiously, it is "self-revealing" (Dainton, 2006, p.57) in the sense that it reveals itself without the need of a distinct awareness or HOR.

Notice that "self-revealing" can here be read in two ways: the thing *doing* the revealing is itself (not some distinct awareness) and the *thing revealed* is itself. As it stands this says nothing of experience revealing itself *to itself*. This third kind of self-revelation *is* entailed, however, when the experience in question is self-conscious in the way already defined. Recall that inner awareness occurs when an experience is revealed to its subject. If one holds that this inner awareness is present *and* that this experience is self-intimating then it follows that the experience is revealing itself *to itself*. The experience is revealed *to* someone or something but, according to the *Simple Conception*, an experience is self-intimating; it is not revealed to a distinct awareness. Thus, it must be the case that it is revealed to the only eligible conscious candidate left: *itself*.

There are two possible meanings of the phrase "consciousness of *e*". The first refers to inner awareness: there is a consciousness of *e* in the sense that someone or something is conscious of *e*. The second refers to self-intimation: in this sense "consciousness of *e*" just means *e*'s *being conscious*, which is to say *e*'s having the property of consciousness; *e*'s being phenomenally present or "revealed". What The Exclusivity Argument claims is that when there is "consciousness of *e*" in the first sense *and e* is self-revealing, *e* is therefore conscious of *e*.

Suppose, for the sake of argument (and contrary to the conclusion of the previous section) that an experience occurs lacking inner awareness. We can ask: who or what is this experience revealed to? The *Simple Conception*, which I take to be the correct view for the reasons we have looked at, claims the following: it is not *revealed to* anyone or anything. It is just *there*. It is self-intimating or self-revealed. It is not made

conscious by falling under the gaze of a distinct awareness. In other words, phenomenal properties are conscious just by virtue of being the kind of properties they are. But now, suppose inner awareness is added to this experience. It becomes true, by definition, that it *is* revealed to someone or something (as this is what inner awareness amounts to: when a subject is aware of its experience, its experience is phenomenally present to that subject). But to who or what is it revealed? Given that *e* is self-intimating it cannot be that it is revealed to or through some distinct consciousness. It must therefore be that it is revealed to itself: there is no other option, i.e. *e* is the *exclusive* candidate. Thus, the subject (the one who is conscious of *e*) must be identical with *e*.

There are a number of ways to reject this conclusion. The argument appears valid, so its soundness is where this must happen. Both clauses of P1 can be attacked. We saw as much in the foregoing sections, but I hope to have shown that we have strong reasons to accept the claim that experience is necessarily self-intimating and self-conscious (in the way defined). With regard to self-consciousness, in this limited sense, it is worth noting that, for many, "it has seemed so obvious as hardly to need defense" (MacKenzie, 2007, p.41). The necessary existence of inner awareness for any given experience can seem to be an *analytical* truth:

> It is impossible to think or experience something consciously without thinking or experiencing it self-consciously, i.e., without being peripherally aware of thinking or experiencing it. (Kriegel, 2004 p.200)

In his own defence of the notion, Strawson (2013, p.8) notes that Aristotle (1936) accepted it too. Sartre, also, is in agreement when he states that "what can properly be called subjectivity is consciousness (of) consciousness" (Sartre, 2010, p.17). I have advanced an argument in support of such a view in §1 (pp.16-39), but it is *tempting* to hold that no such argument is necessary.

We can flesh out The Exclusivity Argument as follows: for any given experience (*e*) for a subject (*s*),

P1. *e* is self-revealing (*e* is not revealed to a distinct awareness)
P2. *e* is revealed to *s* (*s* is conscious of *e*)
P3. an experience cannot be revealed to *x* if *x* lacks any form of awareness

P4. given P3: *e* is not revealed to an *x* lacking any form of awareness

P5. given P2 and P4: *s* cannot be an *x* lacking any form of awareness, i.e. *s* possesses awareness

P6. given P1: *s*'s awareness cannot be distinct from *e*

C. *s* is *e*

Typically we think of ourselves as the kind of beings that *have* experiences: not as the experiences themselves. The Exclusivity Argument gives us reason to believe that we are incorrect in thinking this.

In a nutshell, then, *The Identity Thesis* supports the following line of thought: we do not tell the whole story when we say that a subject of experience is conscious. More accurately, we should say that a subject of experience is consciousness. This prompts some important questions. Switching to the first-person for simplicity's sake, *The Identity Thesis* identifies *what* I am. At the very least, it tells me what *kind* of being I am. As it stands, however, it does not tell me *who* I am; it does not tell me *which* consciousness is "me" except to say that I am the current experience of which I am aware. In other words, it says nothing of the persistence conditions for personhood or of how subjects are to be individuated. These questions need addressing and some crucial objections need to be tackled.

4: The Stream of Consciousness

The Identity Thesis holds that we are, to put it simply, episodes of consciousness. In a nutshell the reasoning is as follows: if you are conscious of an experience (in the immediate, non-inferential way that we typically are)[23] then you must *be* that experience, given that only the self-intimating experience (*e*) can be immediately conscious of *e*. We are not merely "conscious beings" (if this just means "beings who have conscious experiences"); rather, our being (our nature) *is* consciousness. This claim, however, is compatible with very different theories concerning our identity over time. Given that I am defending an experiential account of selfhood, the question of our persistence conditions will necessarily lead us into the infamously difficult terrain of temporal consciousness and its problems. Although a definitive solution to these problems cannot here be offered, a solid account can be defended that is both internally consistent and in accordance with our most basic intuitions concerning our moment-to-moment survival. It also necessitates a reinterpretation of *The Identity Thesis* and provides its successor with a robust framework for individuating selves.

§4.1 Synchronic unity

Before dealing with the relations between experiences across time it is helpful to become clear on the relations between the elements of any given experience in one moment. Happily, a claim can be made here that seems difficult to find fault with: namely that "[from] a phenomenological perspective, it is perfectly obvious that our typical streams of consciousness are unified at any given time" (Dainton, 2004b, p.368). In other words, there exists a synchronic unity within experience. As you read these words the various elements of your experience at any given moment - the look of the page, the sounds you are hearing, the thoughts you are entertaining - are all given *together*. Furthermore, this "unity or togetherness is itself something we

[23] We can, in a sense, be conscious of experiences that are not our own: if I observe someone laughing I can infer that they are in a state of happiness. In a sense I am therefore conscious of their experience. This is clearly not how we are conscious of our own experiences: we do not *infer* their existence - we are immediately conscious of them.

experience" (Dainton, 2004b, p.368). The evidence for these claims is phenomenological: focus on two seemingly distinct aspects of your current experience and notice that, firstly, they are given together and, secondly, that this unity is *itself* something you experience. To the second point, note that "there is something distinctive that it is like to experience these two experiences together, rather than separately" (Dainton, 2004b, p.368). Consider the following description of what it is like to smell coffee and listen to bird-song at the same time:

> There is something it is like to have the auditory experience, there is something it is like to have the olfactory experience, and there is something it is like to have both the auditory and olfactory experiences together. These two experiences occur as parts or components or aspects of a larger, more complex experience. And what holds of these two experiences seems to hold - at least in normal contexts - of all of one's simultaneous experiences: they seem to be subsumed by a single, maximal experience. (Bayne, 2004, p.219)

Dainton calls this ubiquitous relation "co-consciousness" and holds that, so long as we are performing a merely phenomenological investigation (i.e. remaining agnostic on the causal mechanisms involved in the production of consciousness), not much more can be said on this particular matter: such unity is a primitive feature of consciousness and can thus only be analysed up to a certain extent. As Dainton puts it: "[unified] states of consciousness simply consist of experiences related to one another by co-consciousness, and that is all there is to be said" (Dainton, 2004b, p.369). Arguably all theories must bottom out in some kind of primitive given or logical assumption, and the unity of consciousness at a time is one that we can certainly live with: we can verify it any time we care to focus our attention on the togetherness we experience constantly.

David Chalmers and Tim Bayne (2003) have offered an account of the synchronic unity of consciousness that differs, in certain respects, from Dainton's approach. They offer strong arguments in defence of the following Unity Thesis (UT):

> **Unity Thesis**: Necessarily, any set of conscious states of a subject at a time is unified. (Bayne and Chalmers, 2003, p.24)

They provide the following reasonable account of the scope of their aims:

> Our central project will be to isolate a notion of unity on which the unity thesis is both substantive and plausible. That is, we aim to find a more precise version of the unity thesis that is neither trivially true nor obviously false … we aim to suggest at least that the thesis is plausible, that it captures a strong intuition about the nature of consciousness, and that there are no knockdown arguments against it. (Bayne and Chalmers, 2003, p.24)

One way to attempt to account for the unity of consciousness is to appeal to the subject: if we isolate a set of experiences belonging to a certain subject, we can say that those experiences are unified in virtue of belonging to *one* subject. The problem with this approach is that this kind of synchronic unity is not merely primitive but also *trivial* and, as Bayne and Chalmers point out, as such "it cannot capture the intuition that there is some non-trivial way in which consciousness is unified" (Bayne and Chalmers, 2003, p.26). This desired notion, according to Bayne and Chalmers, is "subsumptive unity" (2003, p.26).

§4.2 Spatial unity

One way of elucidating subsumptive unity is by contrasting it with the other kinds of unity we find in consciousness. Synchronic experiences can sometimes be unified in terms of the *object* one is aware of: this occurs when two or more experiences are directed at the same object and *experienced* as such. This kind of unity, however, is quite clearly not sufficient for the task at hand: in holding a book in my hands my experiences of its shape and colour are object-unified but the distracting sound of an incoming email is clearly not, even if it is experienced simultaneously with the book. There is another kind of unity that we can identify here: although the email chime and book experiences were not object-unified they were, nevertheless, *spatially* so. The sound of the incoming email was not part of the same visual space, but it was experienced as occurring *somewhere* in my immediate environment: it was spatially unified with the book experience by being spatially related to it. The book was here, in my hands, and the email chime came from a metre or so behind it. Spatial unity is

an admittedly pervasive phenomenon: spatial relations "play a privileged role in our conception of what constitutes a unified physical world: we regard objects that are not related in any other way as belonging to the same universe if they are spatially related" (Dainton, 2004c, p.2). Sitting at my desk now I am experiencing a manifold of things that, despite being quite different kinds of things, enjoy spatial relations with one another. The feel of the chair on my body occupies a central location, my desk and computer are visually experienced as being a little way in front of my body along with the sounds of my typing, and a lawn mower can be heard in the distance. The spatial unity that we experience as a result of our unified physical world is a prominent part of our experiences.

However, there are reasons to doubt that spatial unity is *necessary* in order for synchronic unity to be present. Imagine the following thought experiment proposed by Dainton (2004c, p.3). You are put under a general anaesthetic, during which time your brain is removed from your head and placed in a vat with transceivers fitted to your nerve endings. Transceiver links to your body connect your brain to all of your usual bodily inputs except for sight and sound. These inputs will now come from an artificial head that is not connected to your body. You slowly wake up…

At first the transceiver-links are not turned on and, as such, you have no sense of spatial location. Without any sensory or perceptual experience you feel "distinctly disembodied" (Dainton, 2004c, p.3) and wait quietly with your thoughts. Your artificial head is now placed on a mountaintop and its transceiver-links are turned on: you can hear and see again. You immediately "feel yourself "transported" to the mountain environment" (Dainton, 2004c, p.3): you are looking down over a vast valley and you can hear the wind around you. You continue to take in the view for some time when you are suddenly met with darkness and silence again: the transceiver-links to your artificial head have been switched off. Now, the links connecting your brain and body are switched on: you can feel your body again, although you cannot see or hear anything. You move around and discover that your body is underwater. You explore your watery home for a while: swimming and bouncing over the terrain. Suddenly, your body disappears again: you have been disconnected. A moment later you are looking back over the valley from atop the mountain. This switching continues on: for some time you are gazing over the landscape, then you are swimming blind for an hour, and then back again etc. After a while you become "accustomed to perceiving the world from two vantage points; in

fact, you get used to existing (or so it seems) at two different locations" (Dainton, 2004c, p.3). Dainton proposes yet another curious development…

We can suppose that this strange state of affairs is something you volunteer for as part of a philosophical inquiry into the nature of consciousness: the necessary technological advancements occur rapidly over the next couple of decades and, as part of a team of researchers, you attempt to tackle various problems by engaging in such experiments. Your current topic is the synchronic unity of consciousness and as such the following strategy has been devised in order to verify or falsify the spatial unity theory: both sets of inputs will now be turned on simultaneously. Your first set of experiences is confusing to say the least: visually you have the experience of gazing over a landscape yet when you move your body you feel as though you are underwater. After some time, you get used to having two sensory-fields instead of one: Dainton argues that each sensory-field would come to "constitute a distinct phenomenal space" (Dainton, 2004c, p.4). You would no longer have the phenomenological impression that these two different kinds of experiences were spatially related: your bodily experiences would seem to be occurring in one phenomenal space, and your audio-visual ones in another (Dainton, 2004c, p.4). Crucially, despite such a spatial fragmentation your experiences would remain fully co-conscious: you would still experience your visual and bodily content *together*. This, at least, is Dainton's claim and it seems plausible. If this could happen then it seems that the synchronic unity of consciousness is not dependant on spatial unity. The crucial issue, then, is whether or not this *could* happen. We can certainly imagine such a scenario, but the move from conceivability to possibility is not a straightforward one. As we will see in §5.1 the method of thought experimentation has seen serious criticism directed at it - perhaps, then, we should be cautious of concluding too much from the intuitive responses such scenarios cause us to have. The following claims from Dainton, however, seem reasonable:

> This form of imagination may be a wholly unreliable guide to physical possibility, and a less than wholly reliable guide to logical possibility in general, but as a guide to phenomenal possibility - to the logically possible forms and combinations of experience - its deliverances are not to be scorned. After all, imagined experiences are experiences in their own right! It goes without saying that delusions, mistakes, misdescriptions and misjudgments are possible, in this

> domain as in most others. But if a form of experience is clearly and vividly imaginable - if it can be unambiguously instantiated in the imagination - then in the absence of any counter-evidence, we have a good reason for thinking that this form of experience could actually occur. (Dainton, 2004c, p.8)

There is a different problem that can be raised: accepting that such scenarios are both conceivable and a justifiable guide to phenomenal possibility, we might still want to disagree with Dainton's interpretation of the phenomenology at play. Antti Revonsuo (2003) attempts just this and claims that there are good reasons to think that, instead of experiencing two distinct phenomenal spaces at once, the subject would either switch between these two different spaces intermittently or enjoy a kind of spatial superimposition. Such superimposition would result in a "very unusual phenomenal world" (Revonsuo, 2003, p.7), to say the least, but Revonsuo reworks Dainton's thought experiment into a fairly convincing picture. As with before, your body is underwater and your surrogate head is perched on a mountaintop. The two channels of input connecting them to your brain are now activated simultaneously…

You are looking out over a valley from atop the mountain. At the same time you seem to have an *invisible body* and the space around you "seems to be filled with a cold, invisible liquid" (Revonsuo, 2003, p.7). You can move your body around and interact with various invisible objects but your viewing position remains the same: the invisible world seems to be moving in relation to the visible world (Revonsuo, 2003, p.7). As such, there is just *one* phenomenal space: "an audiovisual phenomenal space filled with transparent liquid moving in relation to your invisible body and the surrounding invisible objects" (Revonsuo, 2003, p.7). Imagining such a scenario takes some work but it does seem to be a coherent picture.

The problem for Revonsuo is that in order to defend spatial unity as *necessary* for the synchronic unity of consciousness, he must not only show that such a superimposition is possible but also that either it *must* occur or the subject's awareness *must* switch between the two spatial fields intermittently. In other words, he needs to provide a convincing argument to the effect that it is *impossible* for a subject to enjoy co-conscious experiences that are spatially unrelated. This is much more difficult to achieve. Consider the following extension of the thought experiment by Dainton:

> suppose that your eyes tell you that you are about two feet away from a tree. You reach out to touch it. However, instead of finding the expected bark and branches, your hand collides with what feels like a large slimy boulder. Your initial impression is confirmed by further tactile investigations, as is the total absence of anything that resembles - to your sense of touch - a tree. As you stare at the tree and its surroundings, trying (and failing) to discern some visual trace of the boulder that feels so solid beneath your touch, you see a squirrel approaching. It hesitates at the foot of the tree, but only for a moment, before running straight up the trunk. An hypothesis which until now had seemed quite promising - that the tree is embedded in and surrounded by an invisible boulder - has just been refuted. Squirrels can reach many hard-to-reach places, but they cannot pass through solid rock! (Dainton, 2004c, pp.6-7)

In short "the single-space hypothesis is vulnerable to empirical refutation, the two-space hypothesis is not" (Dainton, 2004c, p.7). As Dainton argues, you may *at first* have the impression that your visual experience and your sense of touch are connected to a single spatial area but as soon as you begin to investigate your surroundings this impression will likely dissolve: the hypothesis that there is only one spatial area simply *does not make sense* given your experiences (Dainton, 2004c, p.7). When it becomes undeniable that you are located in more than one region of space it seems likely that your brain will attempt to provide experiences that are commensurate with this (Dainton, 2004c, p.7). It is true that, having previously spent so much time navigating a spatially unified world we would be constantly, and sometimes successfully, trying to interpret these strange experiences as occurring in one unified space. But it does not seem obvious that we would always manage to do so (even if we were purposefully *trying* to). It also does not seem obvious that, in failing to do so, we must also fail to enjoy co-conscious experience at all. In short then, spatial unity does not seem necessary for our experiences to be unified in the way they usually are.

The inadequacy of spatial unity is also suggested by less exotic scenarios: consider the previous example of being interrupted by the sound of an incoming email chime when looking at a book. The sense of frustration felt upon hearing the chime does not seem easily described as having any spatial relation to the sound (we will stipulate that the chime was sufficiently long in duration that the emotional response occurred when the sound was still present in experience). This is particularly true for more

pervasive, yet subtle, emotional experiences such as a general feeling of malaise that can colour much of one's experience of the world without itself being the explicit focus of attention.

§4.3 Subsumptive unity

It seems, then, that neither spatial unity nor object unity can fully account for the kind of unity we find in experience. Having seen what the subsumptive unity proposed by Bayne and Chalmers is *not* we now need to offer a positive characterisation of it: "two conscious states are *subsumptively unified* when they are both subsumed by a single state of consciousness" (Bayne and Chalmers, 2003, p.27, original italics). Consider all of the visual experiences that you are having right now: on the subsumptive account these are all subsumed by a single state corresponding to your visual field (Bayne and Chalmers, 2003, p.27). Moving further out, as it were, your state of perceptual consciousness subsumes both your visual and auditory experiences and, further out still, "it does not seem unreasonable to suppose that there is a single encompassing state of consciousness that subsumes all of [your] experiences: perceptual, bodily, emotional, cognitive, and any others" (Bayne and Chalmers, 2003, p.27). Elaborating further, Bayne and Chalmers propose the following:

> We can think of this last encompassing state of consciousness, for a given subject, as the subject's *total* conscious state. When it exists, a subject's total conscious state might be thought of as the subject's conscious *field*. It can be thought of as involving at least a conjunction of each of many more specific conscious states: states of perceptual experience, bodily experience, emotional experience, and so on. But what is important, on the unity thesis, is that this total state is not *just* a conjunction of conscious states. It is also a conscious state in its own right. If such a total conscious state exists, it can serve as the "singularity behind the multiplicity" - the single state of consciousness in which all of a subject's states of consciousness are subsumed. (Bayne and Chalmers, 2003, p.27, original italics)

In short, two phenomenal states are subsumptively unified when they have a "conjoint phenomenology": "a phenomenology of having both states at once that subsumes the phenomenology of the individual states" (Bayne and Chalmers, 2003, p.32). There is something it is like for you to taste wine, and there is something it is like for you to hear your favourite song. When these two experiences have a conjoint phenomenology there is also something it is like for you to have them *together*. The two experiences are subsumed into a complex state and it is this that unifies them. Although this phenomenal unity is, as with Dainton's synchronic co-consciousness, a primitive notion it is not trivial in the way that subject unity is. If a subject is in multiple conscious states then, clearly, the subject is in the conjunction of those states: *this* is trivial (Bayne and Chalmers, 2003, p.34). However, it is not trivial that "this conjunction will itself be, or be subsumed by, a phenomenal state … that there will be *something it is like* to be in the conjunctive state" (Bayne and Chalmers, 2003, p.34, original italics). Consider the idea of a subject whose simultaneous phenomenal states are only partially unified: experience *x* is phenomenally unified with experience *y*, *y* is phenomenally unified with experience *z*, but *z* is not phenomenally unified with *x*. As Bayne and Chalmers claim, "[one] might suspect (as we do) that such a scenario is impossible and perhaps incoherent" (Bayne and Chalmers, 2003, p.34). It is arguably impossible to even *imagine* what it would be like to be the subject of such a phenomenology. On the subsumptive view this is, in part, because there would not *be* anything it is like to be that subject, "or at least there is no single something-it-is-like that captures all the phenomenal states of the subject" (Bayne and Chalmers, 2003, p.34). In other words, there is no complex state into which *x*, *y,* and *z* are *subsumed*.

It is not, then, a trivial kind of unity: it makes a difference in a way that subject unity does not. The possibility of partially unified experience is an interesting one that would require more space than is available here in order to be adequately covered. It is enough to say, for now, that experiences of the kind we typically enjoy do indeed seem to be unified in the way that Bayne and Chalmers describe and that, furthermore, their notion of subsumptive unity is informative. Dainton has argued, however, that subsumptive unity is ultimately dependant on co-consciousness:

> If two experiences, E_1 and E_2, are co-conscious they automatically constitute a more complex experience W, and in virtue of this E_1 and E_2 are subsumed in W. And what goes for E_1 and E_2 goes for any collection of experiences that are all

> mutually co-conscious. Looking at it from the other direction, if we know that W is a fully unified experience which subsumes the simpler experiences E_1 and E_2, we also know that E_1 and E_2 are co-conscious. And what goes for E_1 and E_2 also goes for larger collections of experiences. (Dainton, 2006, pp.256-7)

Furthermore, it seems that a subsumptive description will need to make reference to the co-consciousness of the subsumed experiences in order to give a full account of the phenomenology:

> Since the subsumptive relation applies outwith the phenomenal domain, to capture the manner in which [the maximal conscious state subsuming a collection of simpler experiences] M's parts are unified, it won't be enough simply to state that M's lesser parts are subsumed within it, we also need to specify precisely how these parts are *experienced* as unified. And to do this we will need to point out that irrespective of how we divide M into parts, each and every one of the resulting parts is directly co-conscious with every other part: a relationship of immediately experienced togetherness binds all parts of M into a unified ensemble. As soon as we try to spell out what subsumption in the phenomenal domain *is like* as a mode of unity, co-consciousness quickly enters the picture. (Dainton, 2006, p.258, original italics)

Ultimately, then, it seems that Dainton's notion of co-consciousness is something we cannot do without. There is a further problem for the subsumption account: consider a total state (S) consisting of [e_1, e_2, e_3]. On the approach offered by Bayne and Chalmers [e_1], [e_1, e_2], and [e_2, e_3] are subsumed in the larger experience S. What of [e_1, e_2, e_3]? This is not subsumed in a larger state but it *is* phenomenally unified. Subsumption does not seem to account for the unity of the state considered as a whole. One possible response is to claim that subsumption is reflexive, in which case the total state subsumes *itself*. However, this is arguably a somewhat artificial move and alters the usual meaning of the term.[24] Synchronic co-consciousness, then, is responsible for the unity of conscious experience at a time. What of a set of experiences considered over time?

[24] With thanks to Barry Dainton for this point (personal correspondence).

§4.4 Temporal consciousness: the landscape of options

There are good reasons to think that the very same relation of co-consciousness is responsible for the unity of experience considered in this sense as well:

> the diachronic unity of experience is no different, in essentials, from the synchronic: both are the product of co-consciousness. Just as simultaneous experiences such as a thought, a bodily sensation and a visual experience, can be experienced together, so can successive experiences, experiences occurring at different (but not distant) times. (Dainton, 2006, p.113)

On this view, diachronic co-consciousness explains a feature of experience that we are intimately familiar with: its unity and continuity over time. The successive experiences that we enjoy throughout our conscious life seem to flow into each other seamlessly: together they form a stream of consciousness. Coining the phrase, William James put it as follows:

> Consciousness, then, does not appear to itself chopped up in bits. Such words as 'chain' or 'train' do not describe it fitly as it presents itself in the first instance. It is nothing jointed; it flows. A 'river' or a 'stream' are the metaphors by which it is most naturally described. *In talking of it hereafter, let us call it the stream of thought, of consciousness, or of subjective life*. (James, 1890, p.239, original italics)

As with the unity of consciousness, the proposed stream-like nature of the succession of experiences is a phenomenological claim: it is a claim about how our experiences *seem* to us in having them.

There are three possible responses to this claim: the first is to accept it and to further hold that conscious experiences really do flow uninterruptedly as they seem to (putting aside, for now, the question of dreamless sleep or other extended periods of unconsciousness). The second is to accept that although it *seems* continuous, the structure of experience is in fact quite different from how it appears. The third option is to deny the claim from the start and hold that if we pay closer attention to the nature of our own lived experience we will see that in fact it does not even *seem* to be

continuously flowing. Each view would subsequently give very different answers to the question of personal identity when consciousness and selfhood are as intimately related as *The Identity Thesis* states. It is important, then, to ascertain which view is most accurate.

Before delving into the technical and complex views that have been put forward in accounting for temporal consciousness, it is helpful to begin with the familiar and seemingly obvious: we *directly* experience change and persistence. On the obviousness of our ability to directly experience motion, Broad states it straightforwardly:

> There is no doubt that sensible motion and rest are genuine unanalysable properties of visual sensa. I am aware of them as directly as I am aware of the redness of a red patch, and I could no more describe them to anyone who has never sensed them than I could describe the colour of a pillar-box to a man born blind … The only way to find out whether a sensum does or does not have a certain quality is to inspect the sensum itself as carefully as possible. (Broad, 1923, p.287)

Albeit couched in terminology we have not been employing, these observations are (on the whole) convincing. A "careful inspection", then, is what we now need to engage in. That we do, indeed, directly experience movement seems very difficult to disagree with:

> Consider some basic data. If I hold my hand in front of me and rotate it at the wrist, I see this rotation as clearly as I see my fingers: my hand's movement is as much a part of the intrinsic phenomenal content of my experience as its colour, shape or size. Whenever we see movement, our visual experience has a temporal character; the content of such an experience is as much temporal as it is spatial … the succession of thoughts and perceptions is itself something we experience; the succession is not just a succession of experiences, it is a succession within experience. (Dainton, 2006, p.114)

When we are engaged in a phenomenological inquiry into the nature of consciousness, putting aside questions concerning the causal mechanisms involved or

the relationship between mind and matter, we should allow the experiential data to dictate the theory. Building an increasingly abstract theory from solid, familiar phenomenological grounds invests its claims with reliability. In this sense, Dainton's view concerning temporal consciousness is strong: that we directly experience change and succession is a highly intuitive starting claim. Slowly move your hand from the left to the right side of your visual field. Do you, firstly, see it on the left and, then, somewhere further to the right whilst *merely remembering* its being to the left? Do you *infer* that movement has occurred from the fact that your hand is now to the right in conjunction with your memory that it previously wasn't? Not at all: you just directly experience its movement. A theory of temporal consciousness needs to account for this feature of experience, or provide a robust argument for denying it. Given the seeming obviousness of our directly experiencing change and succession, such an argument will need to be a particularly powerful one.

For anyone unfamiliar with the area it might seem strange that such an obvious feature of everyday experience could be the subject of disagreement. Other familiar features of our phenomenology are readily accepted: there may be a good deal of disagreement over the nature of colour, for example, but the fact that we straightforwardly *experience* it is not often denied. Yet the fact that we directly experience change and succession is arguably just as obvious. This contrast in acceptance can seem even more odd when we consider other less vivid, yet usually accepted, features of experience. Think of listening to a constant humming that gradually decreases in volume until it is barely audible. So long as it is, indeed, still audible realists about experience have no trouble accepting that the subject of the experience is immediately conscious of that feature of their phenomenology when attending to it, even though it has become a subtle one. Compare this with the starkly apparent movement of a car driving by you: the movement of the vehicle is a *prominent* part of your current experience. It is much more discernable for you than a dim humming noise. There are, of course, reasons for the lack of universal acceptance concerning the existence of this phenomenal feature. An important one is the following: the ability to be *directly* conscious of change and succession seems to entail that a conscious experience is not confined to the present instant - it can seemingly exist in two different moments at once. This entailment is, for some, reason enough to deny our ability to be directly conscious of change and succession. Thomas Reid encapsulates such a view when he claims 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" (Reid, 1854, p.235).

§4.5 The Augustinian argument

To set up the dichotomy in such a way is, however, to oversimplify. Before fleshing out the disagreements in adequate detail it is helpful to begin with a basic puzzle at the root of the philosophical problem of consciousness over time, hinted at by Reid above. The problem stems from the following line of thought:

> Augustine reasoned that the present has no duration whatsoever: evidently, what is present is neither past nor future; take any temporal interval, and make it as short as you like; not all of this interval can be present, because the initial part of the interval occurs before the later part; since the same reasoning applies for any finite interval, no matter how short, it seems that the present, strictly speaking, must be a durationless interface between past and future, between what was but is no more, and what will be but is not yet. (Dainton, 2006, p.120)

If this were true it would follow that any conscious experience confined to the present has no duration *whatsoever*. Even if we focus only on the issue of experience at-a-time, this alleged state of affairs is difficult to accept: how could there be anything it is like to experience *x* if the experience of *x* is "literally instantaneous" (Dainton, 2006, p.120)? There would, quite factually, not be the *time* for it to register. Things do not get any more plausible for this view when we consider experience over time either: change and succession are precisely the kinds of things that require the passing of time. If our experiences were confined to an instantaneous existence it seems difficult to see how we might ever come to be aware of such things either directly *or* indirectly.

Thankfully, such a model of experience is not one we need to spend time picking apart:

> There's almost universal agreement that we don't experience the present of experience just as a (moving) point or front in time that is itself temporally dimensionless, but as something that has an intrinsically temporally extended phenomenological character. (Strawson, 2009, p.249)

Indeed, the bulk of the debate concerning temporal consciousness does not consist of a disagreement concerning whether or not experience seems to extend over time (as Strawson points out, this is nearly universally accepted) but over the relation between the subjective experience of time and objective time itself. To put it simply, the disagreement concerns the phenomenological mechanism responsible for our subjective experience of time. The point of broad agreement is fixed on what has come to be known as the "specious present". William James characterised it as follows:

> the practically cognized present is no knife-edge, but a saddle-back, with a certain breadth of its own on which we sit perched, and from which we look in two directions into time. The unit of composition of our perception of time is a duration, with a bow and a stern, as it were - a rearward - and a forward-looking end. It is only as parts of this duration-block that the relation of succession of one end to the other is perceived. We do not first feel one end and then feel the other after it, and from the perception of the succession infer an interval of time between, but we seem to feel the interval of time as a whole, with its two ends embedded in it. (James, 1890, pp.609-10)

Dainton also offers an elucidating description:

> Rather than being confined to a durationless instant, our direct awareness extends (or seems to extend) over a brief interval of time, the so-called "specious present". The contents within these brief stream-phases are experienced together; think of what it is like to hear the successive rat-tats of machine gun fire, or to see a shooting star flash through the sky - the streak-like motion is experienced as a whole even though it extends over a short interval of time. (Dainton, 2012, p.187)

Dainton's above bracketed qualification pinpoints the major disagreement: although both sides of the debate agree that we directly experience change (and persistence) and that our awareness therefore *presently* apprehends contents that are stretched out through objective time, not everyone accepts that it extends in the way that it seems to. Exactly how long the specious present lasts is difficult to ascertain, but it seems to be less than a second: click your fingers in quick succession five times. Depending on the speed of your clicking, two or three of these clicks were experienced together (albeit not simultaneously), whereas upon the last click the first one was present for you only as a memory, if at all. It is this brief duration of the experienced togetherness of non-simultaneous content that is the target of the term "specious present". For some, the specious present is in an important sense an *illusion*: consciousness does not straddle a significant portion of time in the way it seems to. In turn, each of the two major sides of the debate has its own account of the structure of temporal consciousness: it is either "retentional" or "extensional".

§4.6 The extensional approach

Dainton offers a concise summary of the major disagreement concerning how to account for the temporality of immediate experience:

> Extensional theorists take a simple and direct approach: they hold that our awareness (or consciousness) itself extends a short distance through ordinary objective time. Retentional theorists take a different tack: they hold that our awareness is confined within momentary (or very brief) episodes of experiencing, but they also hold that these momentary episodes of experiencing seem to extend over temporal intervals, and hence can contain change and persistence. The appearance of temporal depth is a direct consequence of the contents contained in a specious present. (Dainton, 2008a, p.625)

The extensional approach is, as Dainton notes, the simpler approach.[25] The essential model is as follows. Consider hearing the notes C-D-E-F-G-A played on a piano in fairly quick succession. Suppose that your specious presents seem to last the length of two notes every time: C and D are given *together* in your experience, as C flows smoothly into D. We can represent the stretch of experience comprising this specious present as [C-D] and the following as [E-F]. There is something else that needs to be taken into account: you do not *merely* experience C flowing into D, and then E flowing into F. You also experience D flowing into E: your stretch of experience is *continuously* flowing in a uniform fashion - in other words the phenomenal content of D is diachronically co-conscious with that of E. We are in effect missing the specious present [D-E]. The retentional approach has its way of accommodating [D-E] but, as we shall see, faces considerable objections. The simpler approach, available to an extensional theorist and defended by Dainton (2006), is to hold that the stretch of experience comprising specious present [D-E] *overlaps* with its joining experiences. On this account, the stretch of experience comprising your hearing C-D-E-F-G-A is extended through objective time: specious present [C-D] shares its latter part with the beginning of the next specious present [D-E], and so on. We will look more closely at how well such an overlap model accounts for our experience of change shortly, by contrasting it with the problems faced by the retentional approach. For now it is enough to say this: accepting that experience extends through objective time, being comprised of overlapping specious presents, offers an attractively simple account of how we are able to directly perceive change and persistence.

One possible weakness with such an account is that it does not fully explicate the structure of the specious present: it bottoms out in the somewhat impenetrable (perhaps mysterious) notion of diachronic co-consciousness. This objection will only hold sway, however, if we have viable alternative theories that do in fact more aptly elucidate the phenomenological mechanism(s) at play. The retentional approach hopes to do just this but faces a host of problems. If these difficulties appear fatal then the simpler, if seemingly less informative, extensional account should be accepted. The claims of the retentional approach now need investigating.

[25] Dainton is far from alone in making a strong case for this approach: see Hoerl (2009) and Phillips (2010).

§4.7 The retentional approach

A full analysis of the host of retentional accounts on offer is beyond our scope, but some general recurring claims and problems can be identified through tackling the theories of Edmund Husserl, whose work on the topic arguably represents "by far the most sustained attempt to describe and understand temporal awareness in the literature" (Dainton, 2006, p.150). Dainton provides the following general description of Husserl's model:

> A stream of consciousness consists of a compact succession of momentary experiences. Each of these momentary experiences contains a representation of the preceding stretch of the stream. As one momentary experience gives way to another, these representations change in a systematic manner, such that phenomenal items seem to occur in the immediate present and then sink into the past … Each momentary experience comprises a momentary *primal impression* and a simultaneously apprehended sequence of representations, the retentional modifications of preceding primal impressions. (Dainton, 2006, p.151, original italics)

To try and ground this rather abstract picture of things, it is helpful to apply it to an actual experience. Close your eyes for a moment and then quickly flash them open and shut again. Let's focus on the brief moment of experience when you opened your eyes (time *t*): the visual experience of the environment before you was the "primal impression". This had the natural sense of occurring in the present. For Husserl, at *t* there was also a representation of the preceding primal impression of the darkness you experienced when you had your eyes closed. This representation was neither a primal impression (it was not experienced as happening *now*) nor a mere memory: it was a retention and as such was presently experienced as having *just* happened. The same modifications occur in the next moment when you close your eyes again: your visual experience of the scene before you is modified into a retention (and thereby seems to have just happened) within your currently occurring primal impression of darkness which strikes you as occurring *now*.

In short, we are made aware of change and succession despite our experiences only lasting for a very brief (though not instantaneous) amount of time: the mode of

presentation of the various contents of our experience systematically alters and results in the appreciation of currently occurring content fading into the past. As it stands Husserl's model is, although not as simple as an extensional account can be, quite an elegant picture of things. However, the following concern forces further complications: we are not aware only "of the flow of content through our awareness, but we are also aware that our awareness is itself continuous" (Dainton, 2006, p.153). Recognising this, Husserl has to say something like the following with regards to your brief visual experience at time *t*: not only is the visual scene accompanied by a retention of your previous primal impression but that very retention is of the previous primal impression *plus* the retentions that it was harbouring. In other words, when you close your eyes again a moment after *t* it will not merely be the primal impression of the visual scene that will become a retention: the retention that was part of the visual experience at *t* will *also* be included in the retention of *t*'s primal impression. Add to this the fact that the retention at *t* is itself partly comprised of a previous retention and the complexity begins to show:

> A particular primal impression becomes first a retention, then a retention of a retention, then a retention of a retention of a retention, and so on, at each successive stage being conjoined with a new primal impression together with the new and intervening retentions … But individual primal impressions are not retained all by themselves in successive acts; they are retained along with all the retentions … that they were originally apprehended with. This *entire complex* undergoes successive modifications in the succeeding momentary experiences. (Dainton, 2006, p.154, original italics)

Such complexity is not necessarily in itself a problem: the simple experience of briefly opening your eyes and closing them again is presumably produced by an astoundingly complicated set of processes in the brain, despite our ignorance of them (from the first-person perspective). Perhaps it is also in part produced by an accompanying phenomenological mechanism of the complex kind that Husserl proposes - one that we are typically not explicitly aware of. This might be acceptable - given the theoretical gains - if such a mechanism holds up under closer scrutiny. However, Dainton highlights a number of problems facing such a model that do call its accuracy into question.

The first concerns the immediacy of our awareness of change:

> Since primal impressions are momentary, there can be no awareness of change or continuity here, for the familiar reason that a succession of impressions is distinct from an impression of succession. (Dainton, 2006, p.155)

Husserl does not want to claim that primal impressions are extended through any significant amount of time (only that they are not literally instantaneous). This is precisely why he needs to invoke the retentional mechanism in order to account for our awareness of change. The upshot of this however, is that "whatever direct awareness we have of phenomenal duration and continuity is located in the retentional matrix, rather than at the level of primal impression" (Dainton, 2006, p.155). The problem now is the contrast entailed between those things we are made aware of through primal impressions and those that we are only aware of through the retentional matrix. It is not clear, to return to our real world example, that your awareness of change is less direct or immediate than your awareness of the darkness you experience when your eyes are closed. Consider the change from darkness to light. The change itself is, of course, not a visual object that you can hold in your gaze. But it is an immediate and significant part of the overall character of this stretch of experience. Husserl's model requires that there be less immediacy: the character of your visual experience is largely indebted to the primal impression, whereas your awareness of change only comes about owing to the changing status of retentions. This does not seem to accurately reflect the phenomenology: change is as immediately present for us as visual content is.

A more general concern for Husserl's account (and, perhaps, any retentional model) is the intelligibility of retentions themselves. What are they, exactly? How do they do the work afforded them by the theory; how do they present things as having occurred in the "just-past of our current experience" (Dainton, 2006, p.155) rather than merely represent them as having occurred in the past, as memories do? The following charge seems reasonable:

> Husserl is *stipulating* that retentions have precisely the properties they need to have for his purposes. Although they occur in the present, they directly intend the immediate past, the past and nothing else … Husserl tells us what retention

> is not [i.e. either primal impression or memory], and what it does, but provides no explanation as to how it accomplishes this. (Dainton, 2006, pp.155-6, original italics)

Dainton's own account can be attacked in a (somewhat) similar way. The role that co-consciousness plays in accounting for our experience of change and succession can be accused of solving the problem whilst still remaining somewhat mysterious. However, even if this accusation is fair, Dainton's model does not bring in the sort of complexity we see in Husserl's and, on this point, scores points for theoretical parsimony (all things being equal). Furthermore, and more importantly, a retentional framework such as Husserl's carries further serious problems that do not arise for an extensional model. This allows us to judge the two competing views using more concrete criteria than theoretical elegance.

Dainton names these two key problems "*lingering contents* and the *clogging of consciousness*" (Dainton, 2006, p.156). In a nutshell, the lingering contents problem is this:

> According to Husserl … momentary experiences enjoy their moment of full consciousness, then slip away, becoming less and less present before finally fading altogether - only then, after they have left direct awareness altogether, can they appear in the guise of ordinary memory. This does not seem to happen. Contents depart from immediate experience cleanly, leaving no residue, and become immediately accessible to memory. (Dainton, 2006, p.157)

Dainton's point seems both hard to disagree with and seriously damaging to the whole retentional enterprise. Knock once on a hard surface nearby. Is there any phenomenological sense in which the noise you make becomes *gradually* less and less present? Not at all: it is there, and then it isn't. It is immediately present in your awareness and then it is gone, save for a memory that you may or may not choose to attend to. The sound of the knock can be described as becoming *objectively* less and less present, in the sense that its time of occurrence is increasing further into the past as each moment occurs. But this abstraction has little do with the phenomenological claim that Husserl is committed to: the sound of the knock should be becoming less and less *subjectively* present, i.e. gradually less and less present in your awareness.

This description does not fit the data: the knock is phenomenally present for you and then it is nothing but memory. As Dainton remarks, the lingering contents problem "is of an almost embarrassing naivety, given the sophistication of the theories under discussion, but it is certainly serious" (Dainton, 2006, p.157).

The clogging of consciousness issue is also no less serious. Recall that, on Husserl's view, it is not just the foregoing primal impression that is retained, but much more:

> We retain not only the past primal impressions, but our preceding total states of awareness. The latter include not only retentions of the primal impressions which preceded them, but the total states of awareness and their retentional complexes, with these retentions themselves containing retentions of previous total acts and their retentions, and so on. (Dainton, 2006, p.157)

Husserl is stuck on the horn of a dilemma: he must either deny that we are aware of the continuity of our experience, or posit a highly complex network of retentions in order to account for such an awareness. Either option lands him in phenomenologically suspect waters: we *are* aware of the continuity of our experience but we are quite clearly *not* aware of such a dense network of content within our phenomenology. The deeper issue is obvious: we are meant to be conducting a phenomenological investigation into the nature of continuity within consciousness, but retentional models such as Husserl's seem to leave the primary data behind. As Dainton remarks:

> The account Husserl provides of the most elemental feature of consciousness is a purely theoretical construction going far beyond the phenomenological data. The theory seeks to explain how our experience is possible, but it does so by appealing to forms of experience which do not seem to exist. (Dainton, 2006, pp.158-9)

§4.8 Continuity questioned

A simpler and truly experience-based approach such as Dainton's extensional account is much harder to accuse of such a move. However, Galen Strawson's account of consciousness through time seeks to do just this when he argues that in fact consciousness does not even *seem* continuous, if we look carefully. He claims that when he is alone and thinking "I find that my fundamental experience of consciousness is one of *repeated returns into consciousness from a state of complete, if momentary, unconsciousness*" (Strawson, 1999a, p.18, original italics). To clarify:

> Even if it's true that an experiential episode always prompts or conditions its successor in some way, it certainly doesn't follow that there's always some sort of experienced sense of connection, conditioning, continuity, or flow. On the contrary. Sometimes the experience is one of a complete break, an inklingless cut. (Strawson, 2009, p.233)

Strawson's first claim seems right: there does not seem to be a *necessary* entailment of experienced continuity from prior prompting or conditioning. His second point also seems reasonable when considering the flow of thought, which is often erratic: "thought has rather little natural phenomenological continuity or experiential flow, if mine is anything to go by" (Strawson, 2009, p.234). The processes of thinking through a problem, of conceptualising some state of affairs or of narrating our experience to ourselves seem often to involve "inklingless cuts" in a sense: moments where our train of thought is stopped in its tracks, replaced by another (related or not) concern. Dainton agrees but adds the following:

> The observation that our thinking is usually fragmented, full of detours and dead-ends, is quite compatible with the claim that there is continuity elsewhere, most notably in perception, mood and bodily feeling, which together constitute the bulk of our experience. (Dainton, 2006, p.118)

In order for Strawson to secure the experience of a "complete break" in consciousness he needs to claim that more than mere *thought* is discontinuous: the phenomenal background ought to be disrupted. This is a much harder claim to defend. Moving,

ironically quite seamlessly, from thought alone to experience considered as a whole Strawson claims that "a positive sense of complete if momentary absence is often part of the phenomenology" (Strawson, 2009, p.238) and fleshes this out as follows:

> The situation is best described, it seems to me, by saying that it is as if consciousness as a whole is continually *starting* or *restarting*. The basic experience of consciousness is not that there is continuous flowing consciousness subject to various small vicissitudes of apparent disconnection, lapses and doglegs and hiatuses. The basic experience, however much it is smoothed out of attentional awareness in everyday life is, I propose, one of tightly packed but non-seamless series of radically disjunct episodes. The process of consciousness keeps bursting silently out of nothingness, even as it maintains strong contentual continuity from burst to burst, as it so often does … (Strawson, 2009, p.238, original italics)

As is often the case with Strawson, his re-assessment of a long-standing view is both excitingly disorientating and honest: he accepts that perhaps his own consciousness is unusual or that he may be confusing occasional introspectively caused discontinuity with the more ubiquitous kind. To the latter charge he concedes that he has "no reply to the objection that I may just be wrong about this" (Strawson, 2009, p.239). It is obviously not sufficient, however, to merely claim that Strawson may be mistaken. If Strawson *is* able to account for why experience might seem continuous to some people (when in fact it isn't), then he is in a strong position: unlike a defender of continuity Strawson would be able to take the phenomenological claims of *both* sides at face value and offer an explanation for the error of his opponents. Thus, Strawson's version of the retentional approach now needs considering.

Recall that in the retentional approach:

> phenomenal unity (or co-consciousness) is confined to the simultaneous contents possessed by momentary episodes of experience, i.e. the contents of individual Retentional specious presents; contents in different specious presents are never unified in this way … The Retentionalists' confinement of co-consciousness to momentary states inevitably fragments our streams of

> consciousness - for, in effect, each specious present is an entirely discrete episode of experiencing in its own right. (Dainton, 2012, p.204)

This picture of things, if accurate, provides Strawson with a powerful reason for his claim that - for some people at least - consciousness is not experienced as fundamentally continuous. It is *composed* of discrete episodes: why shouldn't this fact be available to careful phenomenological inspection? However, he cannot have it both ways: if we are to give credence to his claim that careful phenomenological inspection, for him, reveals "inklingless cuts" in his consciousness, then he needs to tackle the opposing - orthodox - view that consciousness can seem continuous both typically *and* when we are engaged in a phenomenological investigation. The first part of this opposing claim is perhaps relatively easy for him to deal with: it can seem continuous because a good deal of what we are conscious of (i.e. our surrounding environment) is continuous, which thereby flavours the character of our experience in such a manner. But even if this line of thought can be successfully developed, the latter part of the claim remains: for many of us, when we purposefully engage in phenomenological inquiry we are met with a seeming continuity that cannot so easily be dismissed. If our consciousness really is composed of very brief periods of experience, self-contained and not experientially connected to their neighbours directly, then why does experience seem continuous in the way it does upon examination? Strawson has a somewhat ingenious response to this question.

It will help to ground this discussion in a concrete example. Imagine hearing someone sing the familiar refrain "Do-Re-Mi-Fa-So-La-Ti-Do" in ascending notes with no silence in between words. For the sake of simplicity we will concentrate on your auditory experience alone, and stipulate that your specious presents comprise the length of two syllables every time. Reflect on the fact that when you hear such a sequence each note is experienced as flowing directly into its successor with no break in phenomenal continuity. This is, for most of us, an accurate description of how it seems: the experience of "Do" flows seamlessly into "Re" which then flows seamlessly into "Mi". There is not a hint of disruption or discontinuity in the on-going flow of experience - despite the recognisable beginning of a new syllable each time. An extensional approach such as Dainton's provides a straightforward reason for this impression: our consciousness really *is* flowing uninterruptedly owing to diachronic co-consciousness. More fully: it is flowing uninterruptedly - and experienced as such

- owing to the overlapping of specious presents whose content is phenomenally unified by diachronic co-consciousness. Take the first specious present in this sequence: your auditory experience of "Do-Re". "Re" is heard together, albeit not simultaneously, with "Do". Dainton can now hold, owing to his account being of the extensional kind, that the tail-end of this first specious present overlaps with the beginning of the next specious present "Re-Mi". Given that "Re" is diachronically co-conscious with "Mi" we now have unbroken phenomenal continuity stretching for longer than a single specious present. Strawson cannot make this move: his "units" of experience are self-contained pulses which raises the following problem:

> If diachronic phenomenal unity is confined to the contents within individual specious presents, and never bridges the gap between distinct specious presents, how could the successive brief phases in a stream be experienced as flowing into the next in the way they seem to be? Doesn't the experiencing of uninterrupted flow - experience of the sort that we enjoy all the time - require phenomenal unity to run from one specious present to the next? (Dainton, 2012, p.205)

Strawson has an answer: there is a duplication of content that, from the first-person perspective, provides the illusion of continuity. To return to our (admittedly simplified) example, Strawson can claim the following: the first specious present covers "Do-Re" and the second "Re-Mi". Although there is no phenomenal link between the two specious presents the subject has the experience of "Do" flowing into "Re" and then "Re" flowing into "Mi" and the illusion of a continuous flow occurs. We have seen this duplication of contents previously in the work of Husserl, but Strawson puts it to further use: it explains why, for some people, consciousness does not even *seem* continuous if studied carefully. It is possible, he claims, to catch a glimpse of such breaks. There is, however, a serious problem for this kind of view. On the difference between the retentionalist and extensionalist interpretation of such sequences Dainton poses the following issue:

> If we view experienced successions solely in terms of types of qualitative content (or phenomenal character), then the two sequences are largely

> equivalent. But if we focus instead solely on the interrelations between *token* experiences, they are quite different. (Dainton, 2012, p.205, original italics)

With this in mind, a serious objection can be raised against Strawson's account:

> If one hears "do-re-mi", isn't it perfectly clear that the "re" which is experienced as following directly on from "do" is the same token experience (or the same instance of auditory content) which one experiences as flowing into "mi"? As far as I can see, few features of our experience are more obvious than this. (Dainton, 2012, p.207)

We do not experience a duplicate of "Re" flowing into "Mi": we experience "Do" flowing into the very same content that then flows into "Mi". It is not that we experience a "Re"-type content flowing from "Do" and into "Mi": we experience *"Re"* flowing from one to the next, i.e. that very same "Re"-token. Furthermore:

> Precisely the same considerations apply to stream-phases which are of different apparent durations, e.g., a third or a quarter or a tenth of the duration of the specious present … again, such successions can only be accommodated in Retentional models in fragmented and abbreviated form: no such succession can extend beyond the confines of an individual specious present … Reflect again on the continuous character of ordinary perceptual experience, of how each brief phase flows into its successor in the same kind of way. To put it simply, the Retentional approach is incompatible with the *homogeneous* character of phenomenal succession. (Dainton, 2012, p.208, original italics)

It seems, once again, that a retentional account cannot do full justice to the basic facts of our phenomenology. As Dainton remarks on Strawson's unusual approach:

> He has correctly appreciated that the Retentional model breaks our streams of consciousness down into innumerable entirely discrete fragments. My guess is that most earlier Retentionalists did not fully appreciate that their conception of temporal experience has this consequence. (Dainton, 2012, p.209)

Dainton's extensional account clearly faces no such difficulty: overlapping specious presents, whose content is glued together by diachronic co-consciousness, neatly explain the flow of consciousness. As Dainton himself notes, "in the debates to come the Extensional approach may not hold all the aces, and it may not ultimately prevail, but it does at least start with several distinct and significant advantages" (Dainton, 2008c, p.383). This seems a fair and reasonable evaluation of the situation. There is also another advantage not yet discussed: Dainton's account of temporal consciousness is in line with the intuitive judgements we are inclined to make when thinking through thought experiments concerning personal persistence.

§4.9 From temporal experience to personal persistence

A good deal has been written on the topic of personal identity and thought experiments. Part of the literature consists in an attack on the very method of thought experimentation, within the topic of personal identity but also in philosophy at large.[26] The concern is succinctly summarised by Dan Zahavi when he asks us to consider if "our imagination [is] always trustworthy, does it always attest to metaphysical possibility, or might it occasionally reflect nothing but our own ignorance?" (Zahavi, 2005a, p.4). Our inability to answer him with any strong certainty has led some to hold that such a mode of inquiry is best used merely as an auxiliary method for use in scientific discoveries (whose claims can then go on to be verified or falsified):

> Being by nature a speculative enterprise, philosophy benefits from non-speculative input, such as empirical facts and theories. Science, on the other hand, being testable and less speculative, seems to benefit from speculations such as thought experiments … since philosophy is speculative by its very nature, one should not make it more speculative by concocting recherché thought experiments. On the contrary, one should try to find an antidote: try to make philosophy more empirical, for instance. (Peijnenburg and Atkinson, 2003, p.318)

[26] See Wilkes (1988) for a particularly spirited attack on the reliability of thought experiments in the context of personal identity.

The concern behind such a position is understandable and, without wading into the grand task of determining what exactly philosophy should or should not be, it is not difficult to agree that empirical facts can serve a philosophical theory better than thought experiments. The problem, however, is that in certain areas of inquiry it is not clear which empirical facts will be of relevance until a certain amount of philosophy has already been done. In areas such as this it would be questionable to deny oneself the use of *all* of the tools at one's disposal: conceptual analysis, phenomenological investigation and, sometimes, thought experimentation. The problem of personal identity seems to be exactly such an area. It would certainly be rash to rely entirely on the intuitions triggered by thought experiments, but it would perhaps be equally unwise to ignore them altogether (particularly if they are very powerful intuitions). Dainton illustrates this balanced approach when he claims that "thought experiments have a legitimate but limited role to play in our investigations into our own nature" (Dainton, 2008b, xviii). This legitimacy has two aspects: firstly, the believability of the theory and, secondly, our privileged perspective on the subject matter under investigation. To the first aspect Dainton argues the following:

> A typical human life is woven from several different strands, some organic, some mental. Even if in reality these modes of continuity are not separable, by considering imaginary cases in which they do come apart we may be able to learn something about those elements of our lives we regard as most essential to our survival. If we can effortlessly envisage surviving a procedure which eliminates one specific continuity, we have grounds for supposing that this continuity will not feature in a readily acceptable account of what our persistence requires. If, on the other hand, we find it impossible to imagine ourselves surviving such a rupture, we have grounds for supposing the opposite. If we want a believable account of our persistence conditions, an account in which we can recognize ourselves, then it would surely be foolhardy to ignore evidence such as this. (Dainton, 2008b, p.4)

It is hard to find fault with these claims. If a seemingly empirically supported theory identified a person's most essential self with their left earlobe we would, quite rightly, find it hard (perhaps impossible) to take its conclusion seriously. There are of course theories whose empirical support trumps their counterintuitive claims: it does not

matter if we can no longer clearly recognise the concept of "time" in the context of relativity theory. We allow the evidence to dictate our beliefs. With the issue of selfhood, however, things are not quite so straightforward: revisionary theories in this domain are acceptable but become less convincing the more they redescribe our nature in such a way as to make it *unrecognisable* to us. Just why we should give credence to believability in this domain above others has to do with our privileged position: "[we] are, after all, entities of an unusual kind: we are beings with consciousness and self-consciousness, and as such we have a unique perspective on what our existence and persistence involves" (Dainton, 2008b, p.4). Dainton is quick to acknowledge our fallibility in this respect in noting that, for one thing, introspection does not even inform us of the existence of the brain, but argues that such "inside knowledge" is valuable and "inevitably going to inform our intuitive responses" (Dainton, 2008b, p.4). Such a balanced approach to the issue seems reasonable. At the very least we can say this much: if *The Identity Thesis* finds its claims regarding consciousness over time to be in tune with thought experiment-induced intuitions, then this stands in its favour given our status as self-conscious beings. We are now in a position to consider what *The Identity Thesis* should say concerning the nature of the self over time. Having followed Dainton so far, with regards to the *Simple Conception* of experience, the nature of temporal consciousness and the legitimacy of thought experiments we will soon be compelled to depart from him significantly.

5: The Ephemeral Self

§5.1 Intuition pumps

The primary aim of thought experiments within the literature on the problem of personal identity is to identify which kind (or kinds) of continuity is necessary and sufficient for the continued existence of a self. Broken down into its bare essentials the method is simple: a scenario is devised in which one (or more) of our typical continuities is disrupted and the intuitions provoked by considering the case are used as evidence for or against a particular theory of personal identity. In a nutshell: if we can easily and clearly imagine surviving the disruption of a particular kind of continuity then we have good reason to believe that this kind of continuity is not essential to our continued existence, i.e. that it is not part of our essential nature. Thought experiments of this kind are "intuition pumps" and, often, openly so: they seek to prod our intuitions and take lessons from the results. Dennett, coining the term, has argued that such a method is seriously unreliable and open to significant misuse:

> The most influential thought experiments in recent philosophy of mind have all involved inviting the audience to imagine some specially contrived or stipulated state of affairs, and then - without properly checking to see if this feat of imagination has actually been accomplished - inviting the audience to "notice" various consequences in the fantasy. These "intuition pumps," as I call them, are often fiendishly clever devices. They deserve their fame if only for their seductiveness. (Dennett, 1993, p.282)

He warns us to be vigilant for "the sleight of hand that misdirects the audience" (Dennett, 1993, p.282) and, in this sense, Dennett's use of the phrase "intuition pump" seems to be targeting only misleading (perhaps purposefully so) thought experiments. However, useful and honestly constructed thought experiments can also quite naturally be described as "intuition pumps". The key, then, is not to avoid such a method altogether but to ensure that it is properly executed. This, according to

Dennett, means ensuring that the proposed feat of imagination "has actually been accomplished". As we will see, one way to ensure that this occurs is to take care to provide *all* of the relevant information. When the thought experiment concerns personal persistence, then, this means that all of the relevant continuities should be considered.

A significant number of the scenarios that have been developed in this field have failed to do just this. Broadly speaking, the most popular theories of personal identity in this context defend either a physical or a psychological account of our persistence conditions. As such, their scenarios specify physical and psychological kinds of continuities and attempt to show which one is essential to our nature, i.e. that the disruption of only one of them necessitates personal death. But, as we have seen, there is a third kind of continuity that might be highly relevant to personal persistence: phenomenal continuity. As will become clear in fact, there is a strong case to be made for phenomenal continuity as the *most* intuitively compelling guide to personal persistence, when it is actually specified in the description of the imagined scenario (Bayne & Dainton, 2005). Just why this kind of continuity has been so neglected in the literature is an interesting question, to which there may be a host of intermingling answers. One part of the answer will undoubtedly owe itself to the historical progress of the field: the structure of this particular strand of the debate began as physical versus psychological continuity when John Locke (1975) dismissed the usefulness of the soul in accounting for personal identity, and has thus developed along these (increasingly nuanced) lines. To see how the debate is both significantly changed and improved with the introduction of phenomenal continuity, it will be helpful to start with a simple thought experiment.

Suppose that right at this instant, for some unknown reason, both yourself and the current President of the United States become unconscious and that, furthermore, during this period of unconsciousness your entire psychological nature (your memories, your beliefs, your dispositions, your preferences etc.) is swapped with the psychological nature of the President. Upon gaining consciousness, one person has the memory of having just been reading a thesis on consciousness and selfhood, only to find themselves waking up surrounded by concerned secret service agents and medical staff in the Oval Room. The other person has the memory of having just been engaged in a discussion with the White House Chief of Staff, only to find themselves waking up in a strange room with an unfamiliar document in front of them describing

this very scenario. Which of these two people do you most intuitively identify as being yourself? Whatever the fact of the matter (if, indeed, there is one) the person waking up in the White House will, as is stipulated, *believe* that they are you: they have all of your memories and beliefs.

As it stands, this thought experiment is an intuition pump of the less desirable kind identified by Dennett. For one thing, a sleight of hand occurred in the very first sentence when the scenario was described as a *swapping* of psychological natures. Although this is not as egregious a case of question begging as it would have been if the scenario had been described as a "body-swap", it is still guilty of skirting over an important issue: intuitively, a future psychological state belongs to me if it is connected to my past psychological states in the right way. Normally, this means being causally connected by being instantiated in the same brain. In the swapping scenario this typical causal picture is absent. If we add this consideration into the scenario it becomes less clear how relevant the memories and beliefs of the person waking up in the White House are. Although they would be similar to your previous memories and beliefs they would also be very different: they would not have been connected to your previous ones in the *usual* causal way. This muddies the waters considerably in terms of the intuitive support a "body-swap" scenario can demand.

Paraphrasing Dennett, there is a further problem with the scenario as described: has the invited feat of imagination actually been accomplished? Have you really imagined that your *entire* psychological nature has been swapped (or duplicated)? When you entertain the stipulation that your entire psychological nature is now instantiated in the person waking up in the White House, how do you do so? Probably, you have merely imagined that *you* have woken up in the White House: you have conceptualised the scenario from the first-person perspective as shorthand for imagining that your entire psychology has been duplicated. If this is how you thought through the scenario then it would not be surprising if you were to conclude that, intuitively, the person waking up in the White House is you: this was precisely what you were imagining. This is not to say that employing the first-person perspective in such a domain is to be avoided, only that we need to be careful how we use it.

At present, then, this thought experiment appears to be the kind of intuition pump from which we should avoid drawing conclusions: it does not seem obvious that we are in fact able to clearly imagine the scenario in a non-compromising fashion. Further to this, if we remove the sleight of hand and openly stipulate that the

psychological continuity involved is no longer of the typical kind, causally speaking, then the direction our intuitions should pull us in becomes much less clear. If suddenly, unbeknownst to anyone, the normal causal mechanisms governing reality ceased to exist and were replaced with random occurrences that just so happened to make the world continue on in its seemingly typical way, should we regard our identities as shattered owing to the strange new causal mechanism responsible for our psychological lives etc.? It is not obvious whether we should or not. In short, a thought experiment of the kind detailed above seems to do very little to ease the worries of anyone who is pessimistic as to the usefulness of such a method in this domain. There is a simple way, however, of addressing both of these serious problems: we can redescribe the scenario in such a way that it provokes a powerful and convincing intuitive response without the use of any sleight of hand and, crucially, in a clearly imaginable way. This can be done by introducing the kind of continuity we saw in the previous chapter: phenomenal continuity.

§5.2 Streams and intuitions

In order to let phenomenal continuity do its work we need to alter the sort of scenario we have been considering in one important way, for reasons that will become clear later. This time, suppose that there is no break in consciousness and that, instead, as you are reading these words a blindingly white light appears. You close your eyes but the light is still enveloping you. Your body and the surrounding world have receded into unconsciousness: you cannot feel or hear anything and your thoughts have ceased. You are aware only of the silent light and remain so for some time. Your stream of consciousness before, during and after this period is as it normally is: *continuous*. Each phase of your experience flows smoothly into the next in an unbroken stream. At some point in this meditative state the previously stipulated psychological changes occur immediately and without phenomenal trace.[27] Thoughts now begin to creep back in and you hear some faint noises. You open your eyes. Are you in front of this document or are you in the White House? It is arguably unclear: you have the same criteria to go on as in the previous envisaging of the scenario.

[27] The believability of this stipulation is increased if you consider how few of your innumerable memories and beliefs are present in your current experience.

There is one piece of information, however, that will immediately fix your intuitive response: a description of the visual experience you have upon leaving the meditative state. If it is stipulated that you open your eyes and see this document then, intuitively, it seems that a brainwashing has occurred. If, on the other hand, it is stipulated that you open our eyes and see the inside of the Oval Office then, intuitively, it seems that a body swap has occurred. Crucially, our intuitions follow phenomenal continuity: we feel that we go with our stream of consciousness.

This version of the scenario certainly seems readily imaginable: although unusual events are occurring the set of experiences you are described as undergoing are *easily* conceivable. Furthermore, the unusual causal mechanisms do not produce the *uncertainty* that they did previously. However, Dennett and other like-minded critics would be quick to point out that this version of the scenario is not free from sleight of hand either: in describing the stream and visual experience as *yours* I have begged the question. This, then, needs remedying. Referring back to Dainton's account of temporal consciousness in the previous chapter, consider the following:

> The claim that experiences within a typical stream of consciousness are bound together by purely phenomenal relations has solid phenomenological support. We do not need to look beyond the phenomenal to determine which experiences belong to which streams: simultaneous experiences are 'co-streamal', part of the same stream, only if they are related by synchronic co-consciousness; non-simultaneous experiences are co-streamal only if they are diachronically co-conscious, either directly, i.e., they occur within a single phenomenal present, or indirectly, i.e., they are part of a chain of overlapping specious presents. (Dainton, 2004b, p.379)

As we saw, this account of temporal consciousness faired much better against its rival retentional views. A straightforward implication of the above is that the same subject possesses any two experiences if they are co-streamal. If a subject has the experience *x*, and *x* is diachronically co-conscious with experience *y* (either by occurring within a single specious present or as part of an overlapping chain) then the subject of *x* must also be the subject of *y*.

With this terminology in place we can redescribe the thought experiment in a non-biased way. Consider you current experience, *x*, and the subject of this experience, *s*.

As per the thought experiment, *x* now flows seamlessly into *y*: an experience of white light. This continues to flow on for some time, each phase of whiteness directly connected with the next by diachronic co-consciousness and the overlapping of specious presents. At some point the set of qualitatively identical experiences of light flow into the novel visual experience, *v*, which is had when *s* opens their eyes. It is clear that *v* is had by the same subject or self that previously enjoyed *x*: there was no gap in the intervening chain of diachronically co-conscious specious presents comprising the stream of experiences from *x* through *v*. In other words, if you were the subject of *x* then it follows that you are the subject of *v*. If it is then stipulated that *v* is a visual experience of this document then the scenario appears to be a kind of brainwashing. As strange a scenario as this may be (which sees you staring at this document with confused Presidential thoughts running through your mind) the intuition that you have survived the psychological change is hard to resist if you have seriously engaged with the proposed phenomenology: so long as we stipulate that the subject of *x* remained conscious throughout, it is clear that the self goes with the stream. Dainton puts the point forcefully:

> Try to imagine a scenario in which your stream of consciousness flows on in an ordinary straightforward fashion but *fails to take you with it*. The notion that one could be left behind in this way seems absurd, as absurd as supposing one could cease to exist without ever losing consciousness. Suppose you have been granted the power to experience moving in any way that you desire, at any speed, in any direction. It is easy to imagine oneself finding out what it would be like to zoom backward or forward across time, or to shoot across to the other side of the galaxy in a matter of seconds, or to slide into other dimensions or parallel universes. Can you imagine picking up so much speed that you find out what it is like to leave your stream of consciousness behind? If you think you can imagine what it would be like to do precisely this, are you sure you are not just imagining *extending* your current stream in some unusual way? (Dainton, 2008b, p.26, original italics)

To put it frankly: phenomenal continuity wins the intuition war. So long as it is specified in the thought experiment we can imagine enjoying all kinds of ruptures (be they psychological or physical) whilst remaining in existence. Dainton has

impressively defended the pre-eminence of phenomenal continuity in this area through a series of thought experiments concerning virtual reality. It seems reasonable to predict that their power will only increase in tandem with the virtual reality technologies that are just now beginning to enter public awareness. Given the efficacy of these thought experiments, and the undeniable damage they do to theories of psychological continuity in particular, it will be useful to examine them closely.

Virtual reality technology is currently in what could be described as its infancy: the upcoming competitors in the gaming market essentially consist of special headsets that deliver visual and audio input for its user. Surround sound headphones simulate the virtual sound space, and two screens inside the headset produce images for each eye, rendering a three dimensional world to be looked at. Head-tracking technology brings these aspects together to form a powerful impression: upon hearing a noise behind you, for example, and turning your head around to see its cause, the images are produced in accordance with the movement of the headset and give rise to the impression that you are surveying a real three dimensional environment. In effect you have the sense that you are situated somewhere other than your living room: games designers have remarked that the key to delivering a convincing experience is to instil a sense of "presence" for the gamer, a concept that is entirely novel to the industry. Despite its relative simplicity the effect that such technology can have on its user is powerful. In large part this has to do with the importance of our visual experience in engaging with the world and, happily, this is seemingly one of the easier systems to trick.

It does not seem unreasonable to suppose, however, that - should the technology prove commercially successful and its capabilities continue to improve - we may one day see the arrival of a virtual reality system that bypasses such crude mechanisms that require being strapped to our bodies: information could instead be directly fed into the brain of the participant. Whether or not this will be achievable in reality is difficult to predict owing to its dependence on the progress we make in two domains: in our theorising about the brain's production of consciousness and in our technological achievements. That being said, it does not seem - on the face of it - to be a nomological impossibility, so long as we assume that consciousness arises from a physical process. We will follow Dainton's terminology and call this *virtual reality of the second degree*, or VR-2 (Dainton, 2008b, p.14). If such technology is sufficiently powerful the world presented to the user would seem as real as the one you currently

inhabit. Should a cruel trick be played on someone whereby they were connected to a VR2 program whilst asleep, they would be none the wiser as to the virtual nature of their bedroom upon regaining consciousness. VR2 is, in effect, a science-fiction version of the Cartesian Evil Demon. Although a VR2 user's environment is merely virtual his or her experiences are as real as ours. So too are their memories and beliefs etc. We are currently comparing the strength of phenomenal continuity against its rival psychological accounts and as such we now need to consider what Dainton calls "*virtual reality of the third degree*, or VR-3" (Dainton, 2008b, p.15). In this admittedly much more complex process, VR-3 participants not only have their real environment replaced by an illusory world but also their own psychologies. For an extended but finite period the subject will have all of their own beliefs, memories and personality traits replaced with artificial ones. Quite what such a transformation would be like to undergo is difficult to know, but it is certainly a process we can envisage surviving (so long as it is stipulated that the subject continues to enjoy an unbroken stream of experience throughout). Dainton offers a particularly believable, elucidating and tantalising description of what it might be like to undergo a VR-3 experience, and as such it is worth restating in full:

> By way of an illustration, let us suppose that you have long been an aficionado of VR-2 adventures and you decide to take another trip. Having always been fascinated by the Second World War, you opt to spend a week as a WW2 submarine commander. If you choose the VR-2 option, you retain your current memories and personality traits; if you choose the VR-3 option, your current memories and personality traits are wiped from your brain (and stored in the computer's memory) and your brain is furnished with an entirely different psychology, of the kind a typical WW2 submarine commander might well have had. Since you have already been on several VR-2 vacations, you choose the VR-3 option and its promise of a far more immersive experience. You are connected to the machine, the technician tells you the program is about to begin, and as the echo of his words fades, you find yourself on the windswept deck of a submarine at sea, the freezing salty spray lashing your face as the surprisingly small vessel courses through the heavy sea - there is no trace of your previous environment, or the VR-3 machine. For a few moments you feel surprised by this sudden transition, and the entirely realistic character of the illusion, but then

> a sailor speaks to you, and you find yourself replying, and any sense of passively witnessing a life-like illusion disappears: you are soon completely absorbed by the task in hand, your previous life utterly forgotten. A week later, while in the midst of dealing with a serious depth-charge attack, you find yourself transported from your command post in the bowels of your shuddering submarine to the quiet of what seems to be a warm bed. After a few moments of bafflement, you remember everything. Your original memories, beliefs and personality are back in place, you remember the sudden transition to the deck of the submarine as the program began its run, and everything that has happened over the past week - and remembering the damp, the cold, the terror, the adventure of it all, you decide it was worth every penny. You have not only looked through the eyes of a submarine commander - a VR-2 simulation would provide that - you have experienced what it was like to *be* one. (Dainton, 2008b, pp.15-16, original italics)

The above scenario presents a serious problem for a psychological theory of personal continuity: your psychological traits were wiped from your brain during your adventure. On a psychological account either you did not exist during the virtual reality experience or you were in some sense only alive as the psychological information being held in the memory banks of a computer. Either option seems difficult to believe - particularly so if you have successfully imagined the above scenario, given that your stream of consciousness was stipulated as continuous throughout. You were talking to a technician, and then this experience flowed *directly* into a somewhat bewildering experience of being at sea. How could your stream of consciousness have continued on in such a way without *you* continuing on with it?

That there could be such a drastic change in psychology without a corresponding interruption in phenomenal continuity seems highly plausible. Consider how few of your memories and beliefs are currently impinging on your experience in comparison to the *vast* collection from which you can, at any one moment, choose to recall or be guided by. The independence of phenomenal continuity is perhaps at its most evident during meditative stretches of experiences: for many, the purpose of meditation is to quieten the mind - to stop our incessant inner monologue. Allowing all of one's thoughts to dissipate and refraining from actively conceptualising is difficult, but many claim to be able to accomplish it. Such psychologically barren moments give

way to reportedly pleasant *experiences* of peace and calm: there is no reason to think that phenomenal continuity has been disrupted in the slightest.

An influential argument against the use of thought experimentation needs examining: merely *redescribing* such events can alter the intuitive response we feel when imagining them. Bernard Williams (1970) argues that the response elicited from a thought experiment can be altered by changing the narrative perspective from the third person to the first person. This is, quite clearly, an issue that needs dealing with as "it would seem to be a fundamental problem in the design of an experiment if the result of the experiment was totally dependent on the perspective of the observer" (Coleman, 2000, p.63). The results of scientific experiments are supposed to hold independently of the observer's perspective: if the relevant data changed depending on where the observer was located, such an experiment would need re-evaluating. In actual fact, Williams' highlighting of this issue ultimately serves the phenomenal approach. It neither detracts from the usefulness of thought experimentation in this domain, as some have claimed, nor does it work towards strengthening the physical approach, as Williams claimed. To see why this is so, we need to look at the kind of thought experiment that Williams was dealing with.

This was in fact a "body-swap" thought experiment of the kind we detailed earlier: all of your psychological traits are swapped with those of another person. For many, it seems that should this occur then they would go with their psychology: the person waking up in the White House, recall, has all of your memories and beliefs - they will think and behave like you. A convincing story can be told eliciting such an intuition. Williams, however, contends that an equally strong story can be told that does the opposite:

> In Williams reconfiguration of the 'mind swap' experiment (somewhat paraphrased), person A is told that he will be tortured tomorrow. This prospect obviously fills him with dread. Then he is told that he has nothing to fear, for when the time comes for him to be tortured, he will not remember being told about it beforehand, nor indeed will he be able to remember anything that he knows now, he will have total amnesia. Williams states, correctly I believe, that person A would not find this statement reassuring. Then person A is told that when the time comes for him to be tortured, not only will he have total amnesia, but certain changes will be made to his character, and [sic] have new

> ‘memories’ as well, that correspond to those of another fictitious person. Williams suggests that person A would not be reassured by this either. Finally, person A is told that the memories and character traits which will be implanted into his brain will be copied from the mind of another person who is alive today, and that this will be accomplished by means of a machine to which both person A and the ‘memory donor’ will be connected. This still will give person A no reason to not fear the torture, Williams states. (Coleman, 2000, p.61)

As has already been argued, however, an important piece of information has been left out of both narratives: the relations between the experiences of each person. Once we specify *these* details our intuitions become locked in: the person “follows” their stream of experiences (Bayne & Dainton, 2005). This intuition is irresistible so long as it is stipulated that each subject remains conscious throughout the transformation. Crucially this is true *regardless* of the perspective employed: if described from the first person perspective, where *you* experience a flow of experiences, then it seems obvious that your survival is determined by your stream of experiences. However, even if we switch back to the third person perspective of the original narrative and describe the events merely in terms of any given subject undergoing experience *x* flowing into *y* and then *z,* our intuitive response favours phenomenal continuity. This also remains the case even when emotive techniques are employed: if you are told that tomorrow your body will be subjected to horrifying torture but that, prior to this, your stream of consciousness will have been miraculously instantiated in the brain of a peaceful sunbather it seems likely that you will remain free from fear (if you believe both claims). So long as it is made clear that your current stream of experiences will, without rupture, continue on *elsewhere* it is difficult to feel anything other than sympathy for the conscious being who will be inhabiting your body during the torturous acts.

It seems, then, that when phenomenal continuity is specified in the description of the scenario our intuitive responses are left unaffected by a change in perspective. As such, the unreliability of thought experiments cannot be argued for along these lines. Furthermore, such considerations stand in favour of phenomenal continuity: it is the only kind of continuity that remains unscathed by redescription, which gives us reason to think that it is the most reliable guide to personal persistence.

In short, it seems that it is *impossible* to imagine ourselves separating from our currently occurring stream of consciousness. Although it would be unwise to deduce too much from our inability to imagine such an occurrence, it is an incredibly powerful intuition that should not be ignored. As Dainton succinctly puts it: "[for] an account to be maximally plausible, the identity-conferring relationship should be such that if it holds between an earlier self X and a later self Y it is *impossible for us seriously to doubt* that X and Y are one and the same" (Dainton, 2008b, p.24, original italics). This is exactly what we seem to have arrived at: if subject X and subject Y sit at two ends of an uninterrupted stream of consciousness then it is impossible to seriously doubt that X and Y are one and the same self.

§5.3 The Exhaustive Constitution Thesis

These considerations put pressure on *The Identity Thesis*. Recall that we are individuating experiences by character, time and material basis. *The Identity Thesis* holds that the subject of experience *x* is identical with *x*. Experience *x* at t_1 and its neighbouring experience *y* at t_2 are different experiences *even if* they share a common part and - on *The Identity Thesis* - therefore *different* subjects. Yet, as we saw, it is highly plausible to suppose that directly co-conscious experiences belong to the same subject. There are three possible responses to this dilemma should we wish to accept that The Exclusivity Argument makes an accurate claim regarding our nature. I will consider them in order of plausibility from the weakest to the strongest.

The first option is to appeal to the notion of "partial identity". Lewis (1993) and Armstrong (1989) have both appealed to such a concept in other areas, but it might be applicable to our current concerns. Consider a stream of partially overlapping total experiences: e_1-e_2, e_2-e_3, e_3-e_4, e_4-e_5. We want to account for the numerical identity of the subject of e_1-e_2 with the subject of e_2-e_3 whilst remaining committed to the following claims: firstly; every experience necessarily has a subject, secondly; every experience has *only one* subject, thirdly; there is no distinction between an experience and its subject (this is *The Identity Thesis*), and fourthly; co-conscious experiences have *the same* subject. The partial identity account can offer a solution along the following lines:

1. e_1 has as subject, e_2 has a subject etc.
2. e_1 and e_2 are diachronically co-conscious and hence have the same subject: s_1
3. e_2 and e_3 are diachronically co-conscious and hence have the same subject: s_2
4. Given *The Identity Thesis* s_1 is identical with e_1-e_2 and s_2 is identical with e_2-e_3
5. s_1 and s_2 partially overlap, and are therefore *partially identical*

This account allows for the claim that co-conscious experiences have the same subject: e_2 is had by both s_1 and s_2 but, owing to the fact that *both* of them are in part constituted by e_2, they are not entirely distinct: they are *partially* identical. The entire stream of which e_2 is a part consists of a maximal series of partially identical subjects. This view is not entirely dissimilar from Strawson's: our lifespan is significantly shorter than we tend to think it is but we are not, as with Strawson, isolated pulses of experience - we are partially identical with later and earlier subjects in the same stream of experiences.

This is an interesting option but it does not appear particularly convincing in light of the foregoing discussion concerning our survival and its relation to our stream of consciousness. We saw that, if my stream of consciousness continues on uninterruptedly, then I seem to go with it in an *undiluted* fashion. The conscious being at the end of the stream of experiences is not partially identical with me: it *is* me. I survive; I do not "partially survive". At least, this was the intuitive reaction to such thought experiments.

The second possible response is more promising: there is a reasonable sense in which, if diachronic co-consciousness runs right through a series of overlapping specious presents, this stream of experiences can be described as one (rather long) experience. This becomes slightly more palatable to the ear when we consider the intrinsic dynamism of any given experience: it is always flowing from its earlier parts into its later parts, however minutely it is demarcated. An experience is a process: an experienc*ing*. In order to defend this picture we would need to deny that a maximal experience is the totality of directly co-conscious phenomenal content, and extend this description to include *indirectly* co-conscious content, i.e. the entire stream. If the

stream as a whole is *one* experience, and an experience has only one subject, then the problem appears solved.

Dainton himself considers this issue:

> I will regard any experiential component of a stream of consciousness as 'an experience'. A complete momentary cross-section of a stream is an experience, the complete content of a stream over a given interval is an experience, any combination of co-occurring contents within a stream is an experience … A typical stream of consciousness can be divided into particular experiences in many different ways. Although some divisions are more well founded than others, I will not assume that there is any one best way of dividing a given stream into its constituent parts. (Dainton, 2006, p.23)

The question arises, then: why divide it at all? Obviously, one simple answer is that it is *useful* to do so: it allows us to analyse the structure of experience and communicate clearly with each other about particular aspects of a stream of consciousness. Perhaps there are benefits from *not* dividing it, however, and taking a stream of experiences to in fact be "one experience". Tye (2003) argues just this but there are issues with taking such a stance: in terms of the problem of the unity of consciousness, it is not clear just how informative it is. Dainton presents a powerful case against Tye's view that such a "one experience" account dissolves the problem. Tye claims the following:

> the proposed view best accounts for the facts of unity at a time and unity through time. Nothing that we ordinarily say about experience needs to be given up. No large bullets need to be swallowed. The view is clear and simple; and it explains in a compelling way why the problems of unity for experience seem so intractable. Begin with the assumption that there are individual experiences somehow bundled together by a phenomenal unity relation to form an overarching experience and you will find yourself either supposing that phenomenal unity is something unique and basic about which you can say nothing else at all except that it bundles experiences together to form a unified consciousness, or you will join Hume in confessing that the problem is too hard to be solved. The latter course of action at least has the virtue of candor, but the

best strategy, it seems to me, is simply to give up the assumption. (Tye, 2003, p.107)[28]

We have followed Dainton in his defence of such experiential unity: we saw arguments to the effect that co-consciousness is a real (and experienced) connection that exists between our experiences, and found the account highly plausible. For Tye, such a project results in no more than the stipulation that phenomenal unity is "something unique and basic about which you can say nothing else at all" except that it "bundles" experiences together. It is true that co-consciousness is only analysable up to a certain point, but Tye's pessimism seems unwarranted. The following point from Dainton goes some way to addressing this: even if I suppose that my current stream of experience(s) is in actual fact just one experience, it is still composed of parts and, crucially, "even if these parts are not individual experiences, they are nonetheless *unified* in a distinctive way, and the question of what unifies them remains very much alive" (Dainton, 2006, p.253, original italics). Instead of asking how my visual experience is unified with my auditory experience, I can ask how my visual phenomenal region is unified with my auditory phenomenal region (Dainton, 2006, p.253). Ultimately, it seems that the "one experience" view simply *ignores* the problem that co-consciousness attempts to solve. Further to this, it is not clear that it helps with the current issue either. The reasons for this lead to our third option.

Whether or not we identify a brief period in a stream of consciousness (t_1-t_2) as an experience or as a phenomenal region of one long experience, it remains the fact that the experience/region at t_1-t_2 is distinct from the experience/region at t_2-t_3 or t_3-t_4. Another way of putting the issue is this: a stream of experiences (or one long experience) cannot exist *all at once*, i.e. at the same moment in time. It is in the nature of a stream of consciousness to exist over time, and to have temporal phases. Our third option now presents itself.

[28] His reference to Hume concerns the following passage: "In short there are two principles, which I cannot render consistent; nor is it in my power to renounce either of them, viz. *that all our distinct perceptions are distinct existences, and that the mind never perceives any real connexion among distinct existences*. Did our perceptions either inhere in something simple and individual, or did the mind perceive some real connexion among them, there wou'd be no difficulty in the case. For my part, I must plead the privilege of a sceptic, and confess, that this difficulty is too hard for my understanding. I pretend not, however, to pronounce it absolutely insuperable. Others, perhaps, or myself, upon more mature reflections, may discover some hypothesis, that will reconcile those contradictions." (Hume, 2007, p.400, original italics)

Our earlier analysis of the structure of temporal consciousness points the way forward: Tye's desire to identify a stream of experiences as *one* experience is understandable in the context of the unity that such streams possess. Streams of experiences are not *random* collections of experiences: they consist of parts that are fully unified by synchronic and diachronic co-consciousness. Each and every part is connected either directly or indirectly to each and every other part, and as such a stream of experiences is a unified *whole*. Its parts are not all directly co-conscious, owing to the limited range of diachronic co-consciousness, but it is nevertheless a fundamentally unified whole. What, then, is the relationship between these experiential wholes and ourselves? Given the implausibility of the partial identity view, the intimate relationship we have to our experiences as entailed by The Exclusivity Argument, and the highly plausible view that co-conscious experiences belong to the same subject there is only one obvious route to take: we are identical with our streams.

On this view, we are not in fact identical with the experiences that comprise a stream of consciousness, as *The Identity Thesis* holds: these are merely temporal parts (or phases) of the stream-as-a-whole with which we are identical. We are, in effect, adopting the "perdurance" view of persistence, to use the terminology of Lewis (1986). Perdurance accounts hold that material objects are not wholly present at particular points in time: they have temporal stages and are more accurately conceived of as processes, or events, than "things" or substances that move *through* instants of time. In short: we are not the kind of beings that can exist in their totality all at once.

Such an account clearly clashes with our everyday self-conception: I do not have the impression that what I am is not *wholly* present right now. I feel that I am fully present at any given moment of my life, even though I undergo considerable psychological and physical change throughout it. There are three important points to say in response to this: the considerations of §1, §2, and §3 have shown that there are good reasons to think that we *are* fundamentally mistaken in our usual way of thinking about ourselves. It is not surprising, then, that further implications would go against our pre-theoretical notions of the self. This is a revisionist account and, as such, we should *expect* to learn new and surprising facts concerning our identity, if its premises hold.

Secondly, the perdurance view of identity fits well within the context of our current scientifically informed picture of reality. As with most questions of significance there is still an on-going debate in the literature and no clear winner has been crowned, but the currently popular Block View of space-time is a frontrunner. Briefly put, the Block View holds that "now" is not particularly special: all moments of time in our universe (including our own relative past and future) share the same level of reality. As a result, our conception of objects as spatial entities that exist fully at any given moment and move through instants of time is mistaken. Instead, objects are spatiotemporal in nature: they are extended through space *and* time. As such, an object has temporal parts in just the way that it has spatial parts: an object does not exist in its entirety all at one point in space - it stretches over a spatial region. Similarly, an object does not exist in its entirety all at one point in time - it stretches over a temporal region. Taking consciousness seriously, and regarding it as a spatiotemporal existent, means that we should say the same thing regarding a stream of consciousness.

Thirdly, the perdurance view can allow us to be consistent in holding both that there is no ontological gap between a subject and its currently occurring experience *and* that co-conscious experiences are consubjective. This requires a modification of *The Identity Thesis*. The Exclusivity Argument seemed to force our acceptance of a subject (*s*) as identical with its experience (*e*) through the following reasoning:

P1. *e* is self-revealing (*e* is not revealed to a distinct awareness)
P2. *e* is revealed to *s* (*s* is conscious of *e*)
P3. an experience cannot be revealed to *x* if *x* lacks any form of awareness
P4. given P3: *e* is not revealed to an *x* lacking any form of awareness
P5. given P2 and P4: *s* cannot be an *x* lacking any form of awareness, i.e. *s* possesses awareness
P6. given P1: *s*'s awareness cannot be distinct from *e*
C. *s* is *e*

Given that the argument appears valid and that we saw strong reasons to accept P1 and P2 in §1 and §2, we cannot give up P6 and its essential entailment regarding the relationship between a subject and its experience. The conclusion, however, can be reinterpreted in the light of the perdurance account of identity. On this view, the

subject of a total co-conscious experience (*e*) existing at t_1-t_2 is *constituted* by that experience during that interval of time. This is not to say that the subject is *in part* constituted by *e* during t_1-t_2, as would be the case if we took the subject to be the human being considered as a whole: this is not what the The Exclusivity Argument entails. Instead, this means that during t_1-t_2 the subject of *e* is *exhaustively* constituted by *e*: there is no gap between the subject and its experiential episode in that period of time. In place of *The Identity Thesis*, then, we have what I will call *The Exhaustive Constitution Thesis*:

For any total co-conscious experience (*e*) occurring at t_1-t_2 for a subject (*s*):

P1. *e* is self-revealing (*e* is not revealed to a distinct awareness)
P2. *e* is revealed to *s* (*s* is conscious of *e*)
P3. an experience cannot be revealed to *x* if *x* lacks any form of awareness
P4. given P3: *e* is not revealed to an *x* lacking any form of awareness
P5. given P2 and P4: *s* cannot be an *x* lacking any form of awareness, i.e. *s* possesses awareness
P6. given P1: *s*'s awareness cannot be distinct from *e*
C. *s* is exhaustively constituted by *e*

Thus, *The Exhaustive Constitution Thesis* - in conjunction with the previous account of temporal consciousness - brings us to the following conclusion: the subject of experience *x* is identical with the stream of diachronically co-conscious experiences of which *x* is a part. In other words, you are identical with the currently occurring stream of experiences of which you are (non-inferentially) aware. We can call this *The Stream Thesis*. On this view, an experience lasting the length of a specious present constitutes only a brief *phase* of a subject: the maximal series of partially overlapping total experiences constitutes the subject as a whole.

What of our natural impression that we are *fully* present at any given moment? In one sense, this needs to be given up - along with our natural sense that typical material objects are fully present at any given moment. This is forced on us by the perdurance account of identity over time. There is also, however, a sense in which *The Stream Thesis* can do justice to this impression. The brief phase of a subject that exists during t_1-t_2 is present in the way that only experiences can be: it is

phenomenally present. The phenomenal content of your current experience is *fully* present: it is difficult to even imagine what its being partially present would be like. Although at any given moment the subject *considered as a whole* is not fully present (a stream of consciousness is simply not the kind of thing that can exist in this way), the phase of experience constituting the subject at that moment is present in *both* senses of the word: it exists now, in *this* brief period of time, in its entirety and is also experientially *manifest*.[29]

Of the three options available to us in accepting the spirit of *The Identity Thesis*, restructuring it into *The Exhaustive Constitution Thesis* and accepting *The Stream Thesis* appears by far the strongest: it accepts the highly plausible view that co-conscious experiences are consubjective. We have seen strong arguments in support of such a view and *The Stream Thesis* accepts it without resorting to questionable metaphysical moves: instead of attempting to adopt the strange notion of partial-identity we can commit to the straightforward notion of part-whole identity. Instead of trying to sweep the problem under the carpet simply by labelling a stream of experiences "one experience", we are able to account for the consubjectivity of *different*, but suitably related, experiences. The metaphysical move into the perdurance view is easier to defend: it shares its claims with the widely accepted and empirically supported Block Theory of space-time. It is not, then, a terminological sleight of hand. There is, however, an arguably *more* fundamental problem that arises if we adopt *The Stream Thesis* and it is therefore important to deal with it in detail.

§5.4 The Bridge Problem

We have, then, an account of our identity and persistence conditions. There is more to be said, on the relationship between the human being we typically take ourselves to be and the stream of consciousness that we really are, for example, but a basic picture has emerged. Selves (of our kind) are self-conscious, self-intimating streams of experience. An immediate and obvious problem arises: what of the gaps that do occur

[29] It is also worth mentioning that if diachronic co-consciousness was not as limited as it is (i.e. if our specious presents were longer), then larger phases of a stream of consciousness would be phenomenally present in the way described above. In other words, our seeming confinement to brief intervals is not due to a genuine brevity in our lifespans - we exist for as long as our streams do - but to the *range* of the diachronic co-consciousness relation.

in the experiential life of a typical human being? Diachronic co-consciousness and overlapping specious presents guarantee that so long as we remain conscious our streams are free from gaps. But what of periods of unconsciousness, such as the dreamless sleep we undergo each night? Diachronic co-consciousness, either of the direct or indirect kind, cannot reach over these unconscious periods: there is no phenomenal bond between the last section of a stream before dreamless sleep and the first experience upon waking. Two streams of experience that have no phenomenal continuity between them are just that: two numerically distinct streams. Dainton calls this "The Bridge Problem" (Dainton, 2008b, p.27). If *The Stream Thesis* is true, if you are identical with the conscious stream that you are currently (non-inferentially) aware of, then you cannot also be identical with a *different* stream.

On this account dreamless sleep would not merely be the metaphorical death found in poetry, but a literal one: you would cease to exist the moment you became unconscious, never to reappear. This might be tonight when you drift into dreamless sleep, but it might well be sooner than that - if you were to suffer a swift bang to the head or take a nap! We can call this admittedly radical conception *The Ephemeral Self*. On this view, it is an ephemeral self in the standard sense that it can be very short-lived, but typically it is ephemeral in the sense that an ephemeral insect or plant is: it lives for about a day.

I do not wish to claim that such an account has never before even been considered. David Chalmers, for one, mentions this kind of view in a footnote:

> There is a view … on which we … survive during a single stream of consciousness but not when consciousness ceases. On this view, we may … survive from moment to moment but perhaps not from day to day. I do not endorse this view, but I am not entirely unsympathetic with it. (Chalmers, 2010, p.61)

Such a view has not, to my knowledge, been endorsed or defended in any significant sense. I hope to address this in two fundamental ways: firstly, by showing that there is a (perhaps surprisingly) solid line of reasoning that results in such a view, if certain premises are granted. I have begun this task in §1, §2, §3 and §4 but there is more to say on the matter. Secondly, I will seek to defend the position against what is perhaps its most fundamental weakness: its counterintuitive claims concerning personal death.

I hope to show that Chalmers is right in not being "entirely unsympathetic" towards such a view.

§5.5 The "experience" of sleep

Perhaps, however, we have been too quick to lessen our lifespan in such a way: we have, so far, simply *assumed* that dreamless sleep amounts to unconsciousness. As we saw in a previous section this is the default view of many contemporary theories of consciousness:

> Consciousness consists of inner, qualitative, subjective states and processes of sentience or awareness. Consciousness, so defined, begins when we wake in the morning from a dreamless sleep and continues until we fall asleep again, die, go into a coma, or otherwise become "unconscious". (Searle, 2000, p.559)

It is certainly an intuitive position to hold, but there are those who question it. If a good case can be made for the notion that a form of low-level consciousness always accompanies dreamless sleep then *The Stream Thesis* would have the tools to overcome The Bridge Problem: if such a low-level consciousness were of the same continuous kind that we enjoy during our waking hours, then there would be no need of any bridge as there would be no gap to cross. For this reason, it is a possibility worthy of investigation. Ramesh Kumar Sharma (2001) offers an intriguing argument along just these lines. He begins by identifying the undeniable differences between wakeful consciousness and dreamless sleep: "while we are aware of being awake when we are awake, we are *not* aware of being asleep when we are asleep" (Sharma, 2001, p.210). We can think about what it means to be awake when we are in such a state, and we can report this fact to other people: we can do neither with regards to the state of sleep when we are asleep (Sharma, 2001, p.210). Having said as much, he goes on to suggest the following:

> although there is no awareness, during sleep, of *being* asleep, there is perhaps … an *experience* of sleep - an experience, that is, of what sleep is like - that occurs, and can occur, only during sleep. What I wish to say is that although it

> is only in waking that the phenomenon of, say, the meaning of sleep can become the object of conscious thought, this thought about it, irrespective of the theorizing to which it may lead, cannot become a real and significant possibility unless we have had the experience of it when we were actually asleep. (Sharma, 2001, p.210, original italics)

When entering a state of dreamless sleep tonight, if your last typical experience is diachronically co-conscious with your first dreamless sleep experience, then your stream of consciousness would remain in tact. The *kind* of experiences that you undergo would change (from wakeful experience, to dream experience, to deep sleep experience), but there is no obvious reason to think that this would break the usual phenomenal bonds between them. Sharma holds that we are able to think about dreamless sleep in the way we do only because there exists a corresponding *experience* of sleep. Before looking at his arguments for this position, it is helpful to ask just why the orthodox view is what it is. Sharma has his unsympathetic explanation:

> Now, apart from the (near) total loss of awareness of having a body that it seems to involve, the dreamless sleep is sometimes held … also to involve total lack of consciousness of anything whatever. This is maintained on the seemingly "plausible" ground that in rising from sleep we remember nothing of our "experience" during the interval and that, had we really known anything, we would not have failed, as we invariably do, to have any memory impression of that something … The conclusion drawn, then, is that dreamless sleep constitutes a (sort of) discontinuity in an otherwise unbroken conscious life of percepts, images, and thoughts. (Sharma, 2001, p.214)

There is something to Sharma's reluctance to accept this alone as the grounds for supposing that dreamless sleep amounts to unconsciousness: lacking the memory of an alleged experience (*e*) is not necessarily evidence of the non-existence of *e*. You may not now, for example, be able to recall the experience you were undergoing an hour ago but, if you were awake at that time, you were certainly enjoying one. It is also not unusual to have the memory of a previous night's dream triggered by some event or detail we encounter in our subsequent waking hours. Up until the triggering

event we may have been quite confident in our assertion that we had not undergone any experiences at all during that period of sleep: we had no memory of any experiences, and thus no reason to suppose that we had undergone any. Having been triggered to recall a dream, and hence to recall an experience, we would have to admit that our previous lack of memory had been unreliable evidence of the total absence of experience during that time.[30]

Sharma argues that there is in fact a positive reason to deny the orthodox account:

> I want to say that even if the subject is not self-conscious during sleep, as it is when awake, and, further, even if it is not aware of itself or the body as *being* asleep, it would seem that there is yet *some* experience of sleep, the experience, for instance, of the repose and the unalloyed bliss that attends only sleep and that in a way uniquely represents its meaning. To put it through a transcendental sort of argument, there must be experience of sleep if its wakeful recollection, undeniable in our view, is to make any sense. No creative imagination can be called to assistance that can actually picture what sleep is like, nor can its experience be reconstructed on the analogy of any other experience. (Sharma, 2001, p.223, original italics)

We may not, upon waking, be able to recall any explicit content from the period of dreamless sleep but, for Sharma, there is *something* we can recall: having slept peacefully. This, it seems to me, is what he is getting at when he references the "unalloyed bliss" of dreamless sleep. His argument can be put the following way: if dreamless sleep was always accompanied by total unconsciousness then we would not be prone, as we are, to look forward to deep sleep in the way that we do and nor would we be able to report, upon waking, that we had just undergone a peaceful rest. A blissful sleep is something we can look forward to in just the way that we look forward to an enjoyable waking experience. If it did, in fact, represent the total cessation of experience then this would make no sense - or so the argument goes. How could a total lack of consciousness result in a future recollection of that period as blissful?

[30] It also seems to be the case that we can never successfully recall the final wakeful experience of the day before.

This appears to be the essence of Sharma's argument and it is an intriguing speculation. There is an alternative interpretation available, however: upon waking and feeling refreshed we *attribute* this to our previous state of sleep and, although we do not directly recall an experience of peaceful rest, we retroactively describe that period of unconsciousness in experiential terms owing to our current experience. Does it really make "no sense" to do so? This is a particularly difficult problem to answer, owing to the fact that we cannot investigate it *directly* when we are fully conscious and capable of inquiring into the matter with rigorous analytic thought. Perhaps focussing on the phenomenology of the process of *waking up* will elucidate the matter: Evan Thompson (2015) presents a similar argument to Sharma but begins from this different consideration:

> Consider that although deep sleep creates a gap or a rupture in our consciousness, we often feel the gap immediately upon awakening. Our waking sense that we were just asleep and unknowing is not outside knowledge - like the kind we have when we know about someone else's having been asleep; it is inside, first-hand experience. We are aware of the gap in our consciousness from within our consciousness. (Thompson, 2015, p.4)

Thompson's claim that we "feel the gap" is both important and difficult to verify. If it is true that we directly experience the transition from dreamless sleep into waking consciousness, then it follows that we were undergoing an experience of some sort during dreamless sleep: you cannot *directly* experience the transition from state *A* to state *B* if state *A* is completely absent from your awareness. There is an issue to consider here: even *if* the usual process of awakening consists of an experienced transition from a previous state of sleep into wakefulness this would not necessarily entail that dreamless sleep is experiential: it could be that, just prior to awakening, we tend to enter into a different kind of sleep state (as our body and brain begin to approach wakefulness). This concern might be manageable for Thompson: if the gap is still felt when a subject is purposefully awoken by a third party when in the midst of deep sleep, then his point remains. Let's grant that such a rude awakening would indeed be phenomenologically typical and that there would be a sense of "having just been asleep" in just the way there usually is, as Thompson claims. What should we conclude from this impression? This is a difficult question to answer with any

confidence: being a phenomenological observation it requires that we attend to our own experience in order to verify its accuracy. Recalling, now, what it is like to wake up from deep sleep is difficult to do and, even if seemingly achieved, is one step removed from the actual phenomenology: we are *remembering* the event, and thus interpretive faults have the chance to slip into our reasoning. Unfortunately, attending directly to the phenomenology at the moment of its occurrence is also not easy: for one thing, one would need to remember to do it often enough to count as valid evidence and, furthermore, it is not uncommon for our experience of waking up to have a somewhat confused and hazy character. These are not ideal conditions in which to enact a phenomenological investigation.

Returning, then, to the (slightly) less difficult question of our awareness of sleep in a more general sense, Thompson tackles the memory versus inference issue raised by Sharma:

> When you wake up from a dreamless sleep, you are aware of having had a peaceful sleep. You know this directly from memory, so the argument asserts, not from inference. In other words, you do not need to reason, "I feel well rested now, so I must have had a peaceful sleep." Rather, you are immediately aware of having been happily asleep. Memory, however, presupposes the existence of traces that are themselves caused by previous experiences, so in remembering that you slept peacefully, a peaceful feeling must have been experienced. To put the thought another way, the memory report, "I slept peacefully," would not be possible if awareness were altogether absent from deep sleep; but to say that awareness is present in deep sleep is to say that deep sleep is a mode of consciousness. (Thompson, 2015, p.4)

The key issue, then, is whether or not the claim "I slept peacefully" is a memory report. The alternative viewpoint, as we saw above, is that it is actually an inference of some kind: we wake up feeling refreshed, have no memories of the past few hours, and infer that we slept peacefully. It appears that we have arrived at somewhat of a stalemate with no way to conclusively verify either side of the debate. On the one hand it seems reasonable to claim that you cannot infer a lack of consciousness from a lack of memory of such an occurrence (perhaps memories are simply not formed

during periods of deep sleep experience)[31], yet on the other hand it seems questionable to posit a conscious state that we have no *clear* recollection of in order to account for the general sense of being rested upon waking. Unfortunately, it is not even obvious that future empirical discoveries could help definitively solve the issue: suppose a plausible physicalist account of consciousness is eventually constructed whereby the brain activity necessary and sufficient for wakeful conscious experience of our kind is reliably identified. If this brain activity was observed to be absent from a subject during a period of dreamless sleep we may be tempted to conclude that no consciousness was occurring for them at that time. This would arguably be a rash move to make: the theory would only have identified the neural correlates of *typical* wakeful consciousness (presumably with the help of first-person verbal reports) and there is good reason to think that if deep sleep consciousness exists it is likely to be quite different in kind from wakeful consciousness and, as such, might be associated with an altogether different set of processes within the brain.

For Thompson and Sharma the concept of "having slept peacefully" would be impossible for us to entertain if we were entirely unconscious during deep sleep. In an intriguing twist, Johnstone Jr (1973) has argued that it is only if we undergo a period of true unconsciousness that we can have the concept of "consciousness" in the first place:

> A person could not appropriately acknowledge that he was conscious or claim to be conscious unless he knew in the first place what it *meant* to be conscious. Now my argument is that a person who had never slept … could not know the meaning of *either* "consciousness" or "unconsciousness". (Johnstone Jr, 1973, p.74, original italics)

He goes on:

> Let us consider how we might be tempted to proceed if we were confronted with a person who had never slept (or at least could not remember ever having slept), and we wanted to teach him the meaning of either "consciousness" or

[31] Descartes (1986), of course, identified the self with consciousness. He too held that consciousness persists through deep sleep but that memory-failure occurs and, as Hill (2004, p.2) puts it, for Descartes such memory-failure was a result of "the soul withdrawing - so to speak - from the body (and in particular from the brain)".

> "unconsciousness." We might try to point to consciousness as the property common to all his experiences, or as what made these experiences possible. But if he continued having one experience after another, this attempt would not be helpful. If consciousness were what is common to, or what rendered possible, a certain restricted *series* of experiences, we might be able to point to it as the property absent from a different series, but of course a series of experiences from which consciousness is absent is not a series of experiences at all. (Johnstone Jr, 1973, p.74, original italics)

As Johnstone Jr points out, gesturing to an inanimate object and informing the sleepless person that such an object is unconscious would be of little help if he did not already have the concept of "consciousness" to contrast it with. How, though, would falling asleep help him? Johnstone Jr claims the following:

> Suppose it is dark and I am tired. Suddenly it is light, and I am no longer tired. If I feel that there is a gap in the flow of my experience, I may be inclined to frame a hypothesis to account for this gap. One possible hypothesis is that there has been a temporary interruption in the possibility of my experiencing. But consciousness is precisely this possibility of experiencing, and the interruption is unconsciousness … My position, then, is that until a gap occurs in a person's experience he not only cannot *acknowledge* that a state of unconsciousness has occurred in his life, but also he cannot even *conceive* a state of unconsciousness. (Johnstone Jr, 1973, p.75, original italics)

This is a curious claim and, if true, puts pressure on the views advanced by Sharma and Thompson: we clearly *do* have the concept of unconsciousness (they employ it themselves in arguing that it is not what occurs during deep sleep). If we can only have such a concept as a result of noticing a true experiential break in our stream of consciousness, then it follows that deep sleep cannot be experiential in nature. This is far from a definitive solution to the question, but it does suggest one more reason to be sceptical of an account that has no room for genuine gaps in consciousness. It also speaks to the following concern: it seems that transcendental arguments relating to the possibility of obtaining certain concepts can be mounted in defence of both sides of the debate.

In short, the issue is a difficult one to deal with and, unfortunately, it is not obvious which side of the debate is strongest. It seems that we will have to settle for the following line of thought: until more powerful arguments are presented for the alternative we should side with the default view. There is a general naturalistic point in its defense: if consciousness of our kind is produced by processes in our brains, giving organisms with such a property an evolutionary advantage, then the powering down of such activity during sleep would have energy benefits. If conscious experience of our kind creates a survival advantage by allowing an organism to negotiate its world in a superior way, then this feature would arguably be useless when the organism is not conscious of its environment. Having a mechanism whereby such a process completely ceased during certain periods of inactivity would save on energy. This is, of course, far from an entirely compelling argument. For one thing, a movement from full-on waking consciousness into a lower-level form in deep sleep would also save energy, and perhaps there is a reason why naturally evolved brains cannot entirely "switch off". For now, then, we will have to fall back onto Occam's razor: if we accept that there is no experience during deep sleep then we do not need to introduce an entirely new *kind* of consciousness into our ontology. When the question is as difficult to answer as this one such considerations gain weight owing to the lack of obvious solutions. The Bridge Problem stands, then.

§5.6 Conscious capacities

We have seen that there are solid reasons to hold *The Exhaustive Constitution Thesis* and the ensuing *Stream Thesis* to be true. But, perhaps, it comes at too high a cost. We are, after all, seeking an account of selfhood that is both philosophically sound *and* believable. Can you seriously entertain the idea that you did not exist yesterday? As it turns out, there is a solid case to be made for the believability of *The Stream Thesis* or, at least, its *relative* believability. Even further to this, in fact, its implications concerning death, whilst certainly radical, might also be somewhat liberating. Before addressing these issues, it is necessary to see why we cannot follow Dainton in his dealing with The Bridge Problem. Dainton's theories concerning the self-intimation of experience, the temporality of consciousness, and phenomenal continuity as a guide to personal persistence have all been found to be robust and agreeable. It will be

useful to see just why our prior commitments do not allow us to follow him any further.

Dainton identifies two very different conceptions of the self, one of which gives his account the tools to overcome The Bridge Problem. However, as will be shown, moving his theory in this direction brings a new problem that leads to potentially less acceptable claims than those entailed by *The Stream Thesis*. The first conception is the following:

> *The Essentially Conscious Self (ECS):* a self is a thing whose essential nature it is to be conscious; a self is experiencing at every moment at which it exists; a self cannot lose consciousness and continue to exist. (Dainton, 2008b, p.77)

The Exhaustive Constitution Thesis, as we saw, straightforwardly demands that we accept this conception of the self: in saying that you are exhaustively constituted by the currently occurring experience of which you are non-inferentially aware, it is obvious that you cannot exist without the occurrence of an experience. Strictly speaking, the above description does not mirror the claims of *The Stream Thesis* exactly, which holds that the self is a thing whose essential nature it is to be consciousness. The difference between these two ways of phrasing the claim speaks to important concerns that will become clear shortly. Interestingly, the second clause of Dainton's definition can be read in both ways: "a self is experiencing at every moment that it exists" can be taken in the usual sense of meaning that the self is having experiences, but it can also be taken in the sense implied by *The Stream Thesis*: a self *is* experiencing, i.e. a self is identical with a stream of experiential episodes.

The second conception Dainton identifies is the following:

> *The Potentially Conscious Self (PCS):* a self is a thing that is *capable* of being conscious; a self has the *capacity* for consciousness at every moment at which it exists, and it possesses this capacity essentially. A self can lose consciousness provided it retains the potential to be conscious. (Dainton, 2008b, p.79, original italics)

This conception, which Dainton favours, has its obvious advantages. For one thing, it is quite clearly in tune with how we typically see ourselves: we do not usually regard unconsciousness as death - so long as a person retains the capacity to wake up again we regard them as still very much alive. This conception has another advantage: it can deal with The Bridge Problem relatively easily (or so it seems) whilst still paying due respect to our intuitions concerning phenomenal continuity and personal identity. We now need to look at how Dainton develops his account: it is an intricate and detailed theory whose full worth cannot adequately be covered here, but laying out its fundamental basics will provide us with enough to motivate an important dilemma.

For Dainton, the key to overcoming The Bridge Problem is the self's *capacity* for consciousness. Fleshing this out, Dainton states the following:

> I shall call nomologically grounded capacities which produce conscious experience (of any kind) when activated *experiential powers*. I shall further assume that experiential powers are to some degree modular, i.e. that a person's overall capacity for experience at any given moment consists of a variety of different and independent experiential powers … Experiential powers are to be thought of as typically persisting dispositional properties, akin to inertia or electrical charge. (Dainton, 1996, pp.25-6, original italics)

Recall that previously it was shown that two (synchronically or (directly or indirectly) diachronically) co-conscious experiences necessarily belong to the same self or subject. In short, "co-conscious experiences are co-personal" (Dainton, 1996, p.26) or "consubjective" (Dainton, 2008b, p.25). A simple but powerful move can now be made owing to that fact that there "is a natural and compelling way to extend this to experiential powers: by defining the co-personality of experiential powers by reference to the *co-consciousness of their potential manifestations*" (Dainton, 1996, p.26, original italics).

Consider your current experience. All of the experiential powers that are currently producing the various elements of your experience are co-personal in virtue of the fact that they are producing co-conscious (and thus co-personal) experiences. Further to this, there are experiential powers that are not currently active but that, if they were, *would be* producing experiences bound by co-consciousness to the one you are currently enjoying. These dormant powers, then, are also co-personal with your

current experience. If we now consider your ongoing stream of experiences the same fundamental principle can be reapplied: an experiential power (*p*) that is active five minutes from now will be co-personal with the power producing your current experience (*x*) if the experience that it produces is indirectly diachronically co-conscious with *x*. Recall that an experience is indirectly diachronically co-conscious with another experience if between the two there exists an unbroken stream of diachronically co-conscious overlapping specious presents, each glued seamlessly to the next by a direct phenomenal bond. Crucially, we can also say that if *p* is in fact *dormant* five minutes from now it is still co-personal with *x* given that, if it *had* been active, it would have produced an experience that was co-personal with *x* in virtue of the unbroken stream of experiences that *could* have existed between them but, in fact, did not. In short the consubjectivity-conferring co-consciousness of experiences has moved down a level: any two experiential powers are consubjective if the experiences produced by them would have been either directly or indirectly co-conscious had they been active.

Defining the maximal collection of active and dormant co-personal experiential powers as a "C-system", Dainton can now make the following ontological claim and overcome The Bridge Problem: "Selves (or subjects) are C-systems" (Dainton, 2008b, p.113). The forgoing sketch of Dainton's account gives only the barest outline of its essentials: there are multiple ontological options that can be defended in the context of experiential powers and a plethora of general and specific issues that have not even been mentioned here. However, enough has been covered for us to consider the following claims: in order for the C-system theory to adequately overcome The Bridge Problem it must claim that the self or subject *is* the C-system, and this introduces a new problem. To begin with, it is clear that identifying the self with the C-system does indeed overcome The Bridge Problem:

> The move to nomologically grounded potentialities for experience means that gaps between streams of consciousness are no longer in the least problematic. Two temporally separated streams of consciousness are consubjective if they are produced by consubjective experiential powers … Actual experience results from the exercise of experiential powers whose consubjectivity consists in the co-consciousness of their potential manifestations. A particular experiential power persists through periods of quiescence in virtue of the continuous

> experience which would result from its activation. A cycle of being conscious, unconscious, and conscious again is only one of the innumerable forms which the manifestations of the underlying set of consubjective experiential powers could have taken during that period. (Dainton, 2008b, p.114)

It is also clear that the self must be, in some sense, identified *as* the C-system in order for The Bridge Problem to be overcome: *The Exhaustive Constitution Thesis* is perfectly compatible with the notion of a C-system (when the system is merely the experience *producer* and the subject is the stream of experiences) but still faces the issue of gaps in consciousness if the self is identified as the stream of experiences being produced by a particular C-system. The C-system producing my current stream of experiences will tomorrow - after a period of dreamless sleep - produce experiences that *would have been* diachronically co-conscious with my current ones, and hence identical with me. But, as is stipulated, it will not do this: it will produce a new stream of experiences that is not phenomenally connected to my current one, and thus not identical with me. It is only if we identify the self with the C-system that the desired persistence over unconscious periods is secured.

§5.7 Vicarious consciousness

The question, then, is the following: what is wrong with identifying a self (or subject) with its C-system? The short answer is this: a C-system cannot be conscious of its experiences in the way that *we* manifestly are. Or, at least, there are very strong reasons for thinking this.

Consider your own situation. You are currently enjoying a range of experiences that are being produced by your brain. This is, at least, a plausible assumption to make. Taking care not to oversimplify too much, we should note that although your brain has the capacity to produce experiences, it also does much more. It would be a mistake, then, to identify your brain with your C-system. The C-system is that collection of powers instantiated in your brain that produce (or would produce) the co-conscious experiences that you enjoy. With this in mind we can ask the following question: is your C-system conscious of your experiences in the way that you are? If you are identical with your C-system then it ought to be. Recall that, according to the

arguments in favour of *The Exhaustive Constitution Thesis*, you are conscious of your experience whenever it occurs and your experience is self-intimating. The upshot of this, as we saw, was that if you are conscious of an experience (*x*) then you must be exhaustively constituted by *x*, owing to the fact that nothing other than *x* can be aware of *x* in the way the subject of *x* is (i.e. non-inferentially). The problem for the C-system approach is that a collection of experiential powers is not the kind of thing that can be aware of experiences in this way: only experiences *themselves* can be.

Dainton *can* say something like the following: the C-system is conscious in virtue of its self-intimating conscious states, i.e. the C-system *produces* conscious experiences and is in this sense a conscious being. This sense of being "conscious", however, seems to me to be best described as a "vicarious consciousness" and not relevant to the problem at hand. The experiential powers that are currently producing your experience are not themselves conscious of the world or your experience of it. In other words, your experience is not revealed to your C-system: it is "self-revealing", as Dainton puts it (Dainton, 2006, p.57), and given that it is also self-conscious it is revealing itself *only* to itself. We saw this much in detail in §3 (pp.78-81).

That the C-system can be described as conscious in the sense that it is producing experiences fails to address the problem: your current experience is not revealed to your C-system (or the organism of which the C-system is a part) but it *is* revealed to you. To identify the self with the C-system is therefore untenable. The C-system "has" an experience only in the sense that it produces one: but you (the subject of the experience) have experiences in the sense of being immediately aware of them. There is a clear difference between being the "experiential owner" of an experience, and being what we could call the "nomological owner", i.e. the entity whose set of powers are responsible for producing the experience. As was hinted at above this is not just a problem for the C-system approach: any theory that distinguishes our essential selves from our experiences creates a division between the two that cannot then be overcome, if The Exclusivity Argument holds. The subject of your current experience is conscious in the fullest sense of that term: it is immediately and non-inferentially conscious of a phenomenology and thereby conscious of the world. Only a self-intimating experience can be conscious in this way and describing various candidates as conscious merely *in virtue of producing an experience* does not address the problem. It is worth noting that if some form of mind-matter identity theory turns out to be true then it would follow that you are identical with those parts of matter that are

themselves consciousness. It would still not be true to say that you are your brain, however, only that you are identical with those parts of the brain that are themselves the conscious experiences you are currently enjoying, not the parts that are merely producing these experiences.

Something akin to the distinction between "vicarious consciousness" and full-blooded consciousness of the sort we are enjoying right this very moment has been made using different terms, for different reasons, in other areas of the literature. The distinction between "state consciousness" and "creature consciousness" seems to be one such example, and the direction of dependence seems clear:

> Whilst a creature is conscious it may have both conscious and unconscious mental states (and thus it *cannot* be that the conscious status of a mental state simply derives from the conscious status of the creature whose mental state it is). (Manson, 2000, p.407, original italics)

The conscious status of a creature, however, *does* derive from the conscious status of its mental states: we classify a creature as a conscious being if it can have conscious states in general. Again, however, there is an ambiguity in the use of the word "have": the creature can "have" experiences in the sense that part of the creature (its brain or C-system) is producing experiences. But these experiences are not revealed to the creature: they are self-revealing (or self-intimating) and are thus revealed only to themselves (when inner awareness occurs). There is nothing particularly wrong in using the term "conscious" in such a way, but this can become problematic when we are dealing with the fundamental question of our essential nature. The particular way that *we* are conscious of our experiences cannot be applied to a system or an organism (or anything) if such a being is only "conscious" in virtue of *producing* experiences. In short, we are on the inside of consciousness and nothing other than consciousness *can* reach within this sphere of subjectivity. Consider it put in this way: how could an organism or a system become conscious of an experience? It would need to have a conscious state directed at this experience. But, as has been argued, this state itself would be self-intimating and therefore not revealed to the creature. So how could it make the creature conscious of anything *in the way that we are*? It seems it cannot.

The arguments leading to *The Exhaustive Constitution Thesis* clearly have powerful implications that cause problems for more theories of personal identity than

Dainton's. Focusing on Dainton's account has, however, allowed us to appreciate the force of The Bridge Problem. Any experiential account of personal persistence will have to deal with periods of unconsciousness. If the move to C-systems cannot be justified for the reasons outlined here, then we may be forced to accept the strange conception advanced here as *The Ephemeral Self*. We have looked at arguments calling for our philosophical acceptance of it and, although we are far from an *entirely* persuasive account, we have seen a number of phenomenological and analytical considerations in its defense. The next question is this: can such an account also secure our psychological acceptance?

§5.8 Personal identity and believability

It is undeniable that *The Ephemeral Self* is a radical conception of our nature. Despite being grounded in phenomenologically sound observations, and supported by strong intuitions concerning our moment-to-moment survival, its larger implications are strange to say the least. The question then is this: how damaging is this perceived strangeness to the overall viability of the account? A triple-pronged defense can be made: firstly, this is not a problem for this account of selfhood alone. Secondly, *The Ephemeral Self* is an intuitively appealing account in some respects and counterintuitive in others: if the reasons for its intuitive appeal are stronger than the reasons for its perceived strangeness then, on balance, such an account may be more readily acceptable than it appears at first sight. Thirdly, a large part of the oddness of the theory concerns personal death. Such a reconceptualisation, however, may bring benefits that outweigh the potential losses that come with large-scale revisionism such as this. These three strands of the defense are in fact intermingled, but tackling them one by one is a helpful way of elucidating the matter.

The first part of the defense itself comprises two claims: firstly, that this is far from the only account of selfhood that suffers from strange consequences and, secondly, that there may be good reasons to think that *any* account of selfhood will have counterintuitive implications. To the first claim consider the following fundamentally different accounts. Derek Parfit's highly influential theory of personal identity concludes that the question of personal survival can sometimes be *indeterminate*. Parfit describes a scenario in which a scanner encodes a blueprint containing all of his

physical and psychological information, destroys both his brain and his body and then creates an organic Replica from the blueprint. He goes on to make the following points:

> We could say here that my Replica will be me, or we could instead say that he will merely be someone else who is exactly like me. But we should not regard these as competing hypotheses about what will happen. For these to be competing hypotheses, my continued existence must involve a further fact. If my continued existence merely involves physical and psychological continuity, we know just what happens in this case. There will be some future person who will be physically exactly like me, and who will be fully psychologically continuous with me. This psychological continuity will have a reliable cause, the transmission of my blueprint. But this continuity will not have its normal cause, since this future person will not be physically continuous with me. This is a full description of the facts. There is no further fact about which we are ignorant. If personal identity does not involve a further fact, we should not believe that there are two different possibilities: that my Replica will be me, or that he will be someone else who is merely like me. What could make these different possibilities? In what could the difference consist? (Parfit, 1984, p.242)

As we saw earlier, there are *phenomenal* facts that Parfit fails to take into account that can provide such an answer: if the scanning process ends my stream of consciousness (as it certainly seems to in the above scenario) then the Replica is merely similar to me. If my stream of consciousness is (somehow) instantiated in the brain of my clone, with no break in phenomenal continuity, then I have survived (though, of course, not for long). As Chalmers puts it: "I think it is plausible that once one specifies that there is a continuous stream of consciousness over time, there is no longer really an open question about whether one survives" (Chalmers, 2010, p.60). The crucial point at hand, however, concerns the counterintuitive implications of such a theory. Setting aside the philosophical arguments for and against each account, consider which is the more counterintuitive. Is it more far-fetched to suppose that sometimes there may be literally no answer to the question "will I exist tomorrow?" or that you, this stream of consciousness, will indeed cease to exist at some point tonight? It does not seem at all

clear that, on this issue, *The Ephemeral Self* is in any worse a position than Parfit's theory. Evaluating the strength of the two accounts should not, then, be a matter of avoiding the oddest one.

Broadening our scope, it is also not obvious which of the following views is strangest: that we are collections of experience-producing powers, that we are immaterial substances, that we are a "centre of narrative gravity" (Dennett, 1993, p.410), or that there is no such thing as a self and that your name, for example, is no more than a "convenient designation" (Warren, 1957, p.284). There is no shortage of strange notions of the self, and this is the case even when we focus on the phenomenal approach: Bayne, building on Dennett's notion, has argued that the self is best conceived of as a "virtual centre of 'phenomenal gravity'" (Bayne, 2010, p.289). Offering his account as a third way of interpreting the intimate relation between selves and experience, he says the following:

> The two versions of phenomenalism that I have examined each identify selves with concrete particulars: streams of consciousness in the case of naïve phenomenalism and the mechanisms underlying those streams in the case of substrate phenomenalism. But there is another way in which we might hope to forge a constitutive link between streams of consciousness and selves. Rather than looking for something onto which we might map representations of the self, we might think of selves as merely intentional entities - entities whose identity is determined by the cognitive architecture underlying a stream of consciousness. (Bayne, 2010, p.289)

As we have seen *The Ephemeral Self* account is, in Bayne's terminology, a version of naïve phenomenalism. It results in the strange claim that our lives are much shorter than we tend to believe. (Dainton's C-system approach is a version of substrate phenomenalism when it identifies selves with experience-producing powers.) Bayne considers a number of objections to naïve phenomenalism but, on the whole, does not find them to be particularly damaging. His fundamental problem is in fact the following:

> [One] extreme response to the [Bridge Problem] is to bite the bullet, and hold that selves do indeed last only as long as unbroken chains of phenomenal

> continuity allow. This approach might appeal to those who are prepared to hold a radically revisionist conception of the self, but it does not tempt me. My aim here is to understand the role that the self plays within our conceptual scheme, and it is quite clear that we regard ourselves as being able to survive the onset of dreamless sleep. (Bayne, 2010, p.285)

This is certainly the typical way that selves figure in our conceptual scheme. But, firstly, it is not clear that this how we must *necessarily* conceive of ourselves: perhaps we are mistaken and a solid account of how and why we are might lead to a reconceptualisation. Ultimately, the crucial difference at issue here is a methodological one: *The Ephemeral Self* account starts from the phenomenology and allows it to dictate the theory. Bayne, on the other hand, is attempting a conceptual analysis of the role that the self plays. As two different questions there is no conflict. However, Bayne *does* ultimately allow his theory to revise our concept of the self, in which case there is a disagreement concerning what it is that should dictate such revision: phenomenological considerations or our natural conceptual scheme. There is a deeper debate at play here, but it is enough to say this much: it is not obviously unreasonable to lend more weight to the experiential facts - no matter where they lead - than to our pre-theoretical notions of self.

It will be insightful to see just why Bayne's issue with naïve phenomenalism does not stem from the typical objections that can be leveled at it. One objection is ontological: selves are clearly "things in their own right" whereas streams of experiences are not. As we saw, however, co-consciousness addresses this issue:

> Conscious states are not grouped together into streams of consciousness in an arbitrary or haphazard manner. Instead, they are bound together by relations of synchronic and diachronic phenomenal unity. In fact, it is arguable that the forces that knit together the components of a stream of consciousness are no less robust than those that knit together the parts of a single mind or even those that knit together the parts of a single animal. In each case, we have a genuine entity - a thing in its own right. (Bayne, 2010, p.282)

This seems to be the right thing to say. Our pre-theoretical notion of a "stream of consciousness" may well strike us as not being a genuine entity: it seems reasonable

to guess that a large-scale survey of the general public would find such a result. It is tempting to suppose that a significant percentage of people would be naturally disinclined to call a stream of *water* a "thing in its own right", depending on how they interpreted that phrase, let alone a stream of experiences. But, as we saw in §4 (pp.82-108), when we investigate the relations between experiences, both at and over time, the degree of unity found therein is hard to deny. Having done this, it is not difficult to conceive of a stream of experiences as a genuine entity.[32] Strawson has made relevant claims:

> when it comes to deciding which phenomena in the universe count as objects and which do not, there are no good grounds for thinking that non-experiential, non-mental criteria or principles of unity - of the sort that we use to pick out a dog or a chair - are more valid than mental or experiential criteria or principles of unity.
>
> It's arguable, in fact, that there is no more indisputable unity in nature, and therefore no more indisputable physical unity or singularity, and therefore no better candidate for the title 'physical object', than the mental and in particular experiential unity that we come upon when we consider a … subject … in the living moment or lived present of experience. (Strawson, 2009, p.297)

As we saw, Strawson's account of temporal consciousness differs markedly from the model defended here, but his comments regarding experiential unity at a time apply equally well to diachronic unity if we take co-consciousness to be responsible for *both* (and we have seen good reasons to do just this). If unity is what is required for *x* to be considered a "thing in its own right" then a stream of phenomenally unified experiences is a very strong candidate indeed. There is a broader issue at play here: taking experience seriously, a stance that has been adopted from the outset of this investigation, makes it much more difficult to entertain the notion that a series of interconnected experiences is not a genuine entity.[33] Arguments in support of taking

[32] One way of characterising a "genuine entity" or "thing" is to say that it is, in some sense, self-sufficient: it can exist without other things. It is not obvious that a stream of consciousness is a genuine entity in *this* sense. Another plausible option is to hold that a certain level of unity is sufficient for "thinghood". As we have seen: in this context a stream of consciousness is a solid candidate.

[33] Recall that taking experience seriously means to adopt a stance "of robust, full-blooded realism about consciousness. This means taking consciousness as seriously as we take science. From this

experience seriously, then, by necessity support the view that such a diachronically unified stream is a thing in its own right. Such arguments have not been presented here but that there is a convincing case to be made along these lines is a very reasonable background assumption to hold.

A second objection identified by Bayne concerns the sense in which selves can be said to "have" experiences. Identifying selves with streams of experiences seems to entail that a self can "have what it is" or "have itself". Bayne proposes a mereological response: selves have experiences in the way that wholes have parts, i.e. in the way that "universities have philosophy departments" (Bayne, 2010, p.282). As we saw in §3 (pp.78-81) there is an even stronger response available, however: "having an experience" in the sense of being its experiential owner (i.e. the entity to whom it is phenomenally revealed) is a property that can *only* be had by the experience in question. The reasons for thinking this were laid out in The Exclusivity Argument of §3 (pp.78-81). In this sense, naïve pheomenalism (of some form) is in fact the *only* kind of account that can do full justice to the sense in which an experience is owned by a self. Indeed, it was this consideration that forced us to move in this direction to begin with.

A third objection identified by Bayne is that if selves are streams of experiences then most of our everyday self-descriptions are inaccurate: a stream of experiences cannot, for example, drive a car. As Bayne accepts, however, there are ways for a naïve phenomenalist to reinterpret such claims. As we will go on to look at in more detail in §6.1, Strawson's approach - in which he highlights the non-univocal nature of the concept "I" - is one such option. Briefly put: my current stream of experiences is associated with a particular human being and *he* can indeed drive a car.

Although Bayne does not find the above objections to be particularly difficult for a naïve phenomenalist to overcome, his methodological choices mean that he cannot accept such a view: he needs to overcome The Bridge Problem and supplies an account that can do just that. However, as with Dainton's model, this comes at a cost:

> We can now see where other approaches to the self go wrong: they assume that there must be some 'real' entity that plays the role of the self. The only thing that plays the self role - indeed, perhaps the only thing that could play the role

---

perspective, sensory experiences, bodily sensations and conscious thoughts are regarded as just as real as paradigmatic physical things such as mountains, houses and trees…" (Dainton, 2006, p.1).

> of the self - is a merely intentional entity. Experiences do indeed have 'owners' or 'bearers', but the owner of an experience is nothing 'over and above' a virtual object - indeed, the very same virtual object around which that experience is structured … To identify the self with a stream of consciousness or its underlying substrate is a bit like identifying Hercule Poirot with the novels in which he figures. (Bayne, 2010, p.290)

Bayne builds on this metaphor in presenting his solution to The Bridge Problem: we can agree that the same self has different streams of consciousness in just the same way that we can agree that a fictional character is the "same" character across multiple novels: it is, as Bayne admits, ultimately only a matter of convention (Bayne, 2010, p.291). The strength of Bayne's approach depends on his arguments, not on the counterintuitive nature of the idea of such a virtual self. Having said as much, two points are of importance here: firstly, Bayne's intentional approach is one more example of the strange consequences that can result from reasonable premises concerning selfhood and, secondly, his account serves to highlight the difficulty that any experiential theory faces in overcoming gaps in consciousness. Perhaps, then, it is not unreasonable to suppose - as *The Ephemeral Self* account does - that we should bite the bullet with regards to gaps in consciousness: we are simply mistaken when we intuitively feel that we have existed for longer than our current stream of experiences. Is it more believable to suppose that we only survive gaps in our consciousness *conventionally*, in the same way that a fictional character can be the same entity in two different novels? Whether or not it is *you* who exists tomorrow morning certainly doesn't strike us as the kind of question that is answerable only by an appeal to agreed conventions. There seems to be a fact of the matter, no matter how difficult it is to know. This, at any rate, is by far the more intuitive position.

§5.9 Being real

Another way of making the above point is as follows: *The Ephemeral Self* account may conflict with our sense of how often death occurs, but it *does* respect our intuitions concerning the objective (i.e. non-convention-dependent) fact of its occurrence. Mark Johnston (2010) has defended a theory of personal identity that

clashes with *both* intuitions in its claim that we could extend our lifespans merely by conceptualising ourselves differently. As with Bayne, this results from taking selves to be intentional in nature. Johnston begins by highlighting three issues with the typical thought-experiment approach to the problem of personal identity:

> First, the specific necessary conditions on our survival, conditions that are the upshot of our common essence, need not be available to armchair, or "a priori," reflection. It is after all a Lockean point … that our real essence cannot be discovered by attention to our concepts but only by empirical investigation into what is in fact the case. (Johnston, 2010, p.44)

The claims of *The Ephemeral Self* account are admittedly a result of "armchair reflection": the problems discussed have been the kind of problems that *require* a philosophical analysis even when empirical studies are of use. It does not seem obvious, however, that this has been a merely conceptual analysis: in investigating the structure of our own phenomenology we have, in a certain sense, been undertaking an empirical project. We have tried to let our experience *as it is in itself* dictate our route, and have for this reason endeavored - in this particular sphere at least - to look into "what is in fact the case". One of the key claims forwarded by *The Exhaustive Constitution Thesis* is, after all, that it is the *essential nature* of consciousness to be self-conscious and self-intimating - not that our mere *concept* of consciousness necessarily involves such features. We are in fact acquainted with conscious experience in a much more intimate way than we are with the objects of *scientifically* empirical studies, and as such are able to make claims that are not justified solely in terms of our concepts. Johnston has a further worry:

> Second, in the massive core of cases of ordinary survival from day to day, many sources of evidence for personal survival, such as persistent bodily integrity and mental continuity, converge and agree, whereas the whole philosophical charm and supposed utility of the imagined cases in the literature on personal identity lie precisely in teasing these elements apart. The obvious question arises: Might we not have thereby *undermined* our ability to make good judgments about personal identity when considering these very cases? (Johnston, 2010, pp.44-5, original italics)

We have already seen arguments that attempt to deal with this kind of objection in §4.9. Simply put, thought experiments of this kind can be legitimate (owing to our privileged position as self-conscious beings) albeit limited guides to personal persistence. Johnston is not *entirely* against the use of such a method (he employs it himself, albeit to different ends) only wary of relying too much on the intuitions provoked by such thinking. This much we should agree with. His third concern moves us towards his rival account of the self:

> When we take the trouble to look, we do not find much evidence that in tracking objects and persons through time we are actually deploying knowledge of *sufficient* conditions from cross-time identity. Instead, as a matter of empirical fact, it appears likely that nature saves us inferential labor by having us "offload" the question of sufficiency onto the objects and people themselves - if I may put it that way. (Johnston, 2010, p.45, original italics)

He offers a motto for offloading: "I don't know what the (non-trivial) sufficient conditions for identity over time are, but I do know a persisting object when I see one" (Johnston, 2010, p.45). Johnston takes his point to be entailed from the fact that we do not infer that some object is the same over time by considering identity criteria. Instead, we are directly aware of its movement and thereby immediately aware of it *as* a persisting object. As we saw in §4.4 (pp.92-94) there are good reasons to agree with this account of the phenomenology. Johnston claims that this "off-loading" occurs when we are aware of human beings too: they "capture our attention at various times and *over* time, as when we see them moving or hear them talking" (Johnston, 2010, p.46, original italics). We directly perceive their identity over time by directly perceiving their movement, not by a process of inference concerning facts of personal identity over time. So far, there is a good deal to agree with in Johnston's third concern. If we do indeed track objects, including human beings, that are themselves "naturally individuated" (Johnston, 2010, p.46) in such a way, and not by way of an "implicit grasp of "gensidentity" relations" (Johnston, 2010, p.46) then the reliability of our intuitions concerning thought experiments might well be called into question. This issue, however, does not seem to be damaging to the kind of thought experiments we have been dealing with: the persistence conditions specified by *The*

Stream Thesis are themselves based on our *immediate awareness* of continuity. Instead of an awareness of objectual continuity, however, this was an awareness of phenomenal continuity. Johnston, however, cannot follow us in this direction: he does not accept the reality of the self in the same way that the account presented here does.

To appreciate this we need to introduce Johnston's notion of "an arena of presence" (Johnston, 2010, p.139). Consider your current experience and all of the perceptual content that comprises a good deal of it. You are visually aware of various objects, you are hearing various sounds and you can sense your body and feel anything it is currently touching. Johnston makes the following point: these items are not just free-floating objects in a phenomenal space - their modes of presentation are *perspectival* in the sense that they are organised around a particular viewing position. Or, to be precise, they are organised around an *implied* viewing position. There is an "arena of presence" and you experience yourself as being at the centre of it. This seems to be a phenomenologically accurate description of our typical perceptual experience. Having made the previous claims, Johnston hints at the picture to come when he asks us to think of the arena "as a sort of virtual frame or "container" … it is if you like the mind considered as a sort of place, the mental "bed" in which the stream of consciousness flows" (Johnston, 2010, p.140).

Johnston claims that it is "the property of being me at the center of this arena of presence that is the property of being me in the most intimate and important sense" (Johnston, 2010, p.144). If we grant him this, whilst accepting that such an arena is organised around a merely *implied* position, then we must deny the reality of the self. The centre of the arena is, after all, merely an illusion: an intentional object "answering to nothing in the world" (Johnston, 2010, p.145). *The Ephemeral Self* account faces no such problem for the following reasons: the self undergoing experience *x* is exhaustively constituted by *x* during its occurrence. This includes, but is not limited to, the implied perceptual perspective therein. That we can experience ourselves as being at the centre *only* need not be denied: as we shall see, in §6.5 and §6.6, there are good reasons to think that seeing through such an illusion (as opposed to identifying ourselves with it) can be psychologically beneficial. We have seen, then, one more intuitive gain from identifying ourselves with experience: unless we are illusionists about consciousness in general, it blocks this kind of move from experiential entity to *virtual* entity. Part of an experience can be merely intentional (or virtual) in nature, but the experience as a whole is still very much real. There is

another way in which Johnston's account forfeits intuitive appeal as a direct result of denying the reality of the self: it results in claims concerning personal death that are arguably even more radical than those of *The Stream Thesis*.

Johnston presents the issue forcefully when he asks, "[under] what conditions does the very same arena of presence and action continue on? You have no idea, and neither do I" (Johnston, 2010, p.174). This is, in effect, the same issue raised by Bayne's account: if the self is merely intentional then there are no objective facts concerning its identity over time. Just as with the identity of a fictional character appearing in different novels, the matter is solved by convention. To put it bluntly: whether or not *x* is identical with *y* depends on nothing more than the beliefs of the relevant people. Johnston takes this entailment seriously, and offers an intriguing account of personal death in the process. To simplify: if selves are intentional, if our persistence conditions are a matter of convention - akin to fictional characters - then we can alter our nature merely by adopting a different narrative. On this view what counts as personal survival is "not something fixed in a uniform way by our natures as persons" (Johnston, 2010, p.317) but dependent on our "identity-determining dispositions" (Johnston, 2010, p.317). In other words, how we are disposed to identify with future selves determines the facts concerning our survival. If true, then personal death can be avoided merely by believing certain propositions: believing that you survive teletransportation of the kind we saw described by Parfit would result in survival. Of course, such beliefs would need to be caused by a fundamental change in your identity-determining dispositions, and this is not something that Johnston supposes would be an easy task, but in principle his account allows for it. There are further strange consequences identified by Dainton:

> In a similar, but more radical vein, anyone who believes that it is appropriate to identify with *all* future individuals (or at least, those who are deserving of such concern), and who also succeed in adjusting their identity-determining dispositions accordingly, will be justified in believing that they themselves will survive *as* these future individuals, in a quite literal way. Something approximating immortality is thus a genuine possibility for those who are suitably disposed. (Dainton, 2012, p.176, original italics)

Johnston's account is intriguing to say the least, but deeply counterintuitive. Consider the following: if Johnston is correct then I can solve my own "personal Bridge Problem" with ease - I only need to change my fundamental dispositions regarding what it is possible for me to survive, and thereby come to believe that a gap in my consciousness does not result in my death. But, of course, this is *already* what I am fundamentally disposed to believe (or, at least, until the arguments presented here began to erode my certainty). Is it really plausible to hold that there is no Bridge Problem purely on the basis of what we are disposed to believe? The point can be put quite sharply as follows: if the arguments in support of *The Ephemeral Self* are strong enough then, on Johnston's view, its adherents would not merely be accepting that our lifespans are shorter than they had previously thought: they would be *causing* them to become so! *The Ephemeral Self* account has the counterintuitive conclusion that I will cease to exist tonight but conventionalist accounts, such as Johnston's, have implications that seem absurd by contrast. All of these considerations should be kept in mind when we judge the plausibility of *The Ephemeral Self* on the basis of its intuitive appeal or lack thereof.

There are many routes to a denial of the reality of the self. For someone like Thomas Metzinger: "no such things as selves exist in the world: Nobody ever *was* or *had* a self" (Metzinger, 2003, p.1) and "[f]or all scientific and philosophical purposes, the notion of a self - as a theoretical entity - can be safely eliminated" (Metzinger, 2005, p.3). On Metzinger's account what I typically take to be my inner experiential self is in fact a phenomenal model that is not recognised *as* a model by the system producing it:

> We are Ego Machines, natural information-processing systems that arose in the process of biological evolution on this planet … We each have conscious self-models - integrated images of ourselves as a whole, which are fully anchored in background emotions and physical sensations. Therefore, the world simulation constantly being created by our brains is built around a center. But we are unable to experience it as such, or our self-models as models. (Metzinger, 2009, p.207)

Such a denial of the self is also a uniting claim amongst the various strands of Buddhism. The radical nature of such a position is arguably sometimes not fully

appreciated - on the difference between reductive and eliminative views the following should be kept in mind:

> Both theories … reject the notion of a substantive self which somehow exists beyond the bounds of experience. The difference, however, is that while the reductionist accounts then go on to resurrect the self and, consequently, its identity, in terms of putative psychological relations or various theories of the body, the no-self theory lets the self lie where it has fallen. This is because the no-self theory is not a theory about the self at all. It is rather a rejection of all such theories as inherently untenable. (Giles, 1993, p.175)

As we have seen *The Stream Thesis* shortens the lifespan of the self in a counterintuitive way. It does, however, accept its reality to begin with! In short, it seems fair to say this at least: if *The Ephemeral Self* faces a problem in the way it clashes with our everyday intuitions concerning our own nature, it is far from alone in this. There is another point to note: Bayne, Johnston, Metzinger, Dainton and the Buddhists all appeal to *experiential* considerations in reaching their conclusions. By denying the identity of a persisting self with its experiences, they are led to the various odd consequences highlighted above. Accepting such an identity admittedly leads to strange claims too, but it is not obvious that they are more implausible than the alternatives. When considering experience-centered accounts, then, it seems that a purely experiential picture of our nature is an option that we should not discount.

§5.10 Animal instincts

Perhaps a more commonsense account of personal identity, however, might avoid such issues. One such approach is zoological. On this view human beings are a kind of animal and, as such, *you* are one particular animal (the one reading this now). Might it be this simple? Olson (1998, p.654) puts it straightforwardly: “Why couldn’t we be human beings?” A seemingly very simple thought experiment puts pressure on such a view: suppose that you and I swap heads. (As strange as it may seem head (and nervous system) transplants may not be science fiction for much longer, depending on which experts you believe.) The problem for Animalism is brought out by Dennett’s

quip that nobody would opt to be the body-donor in a one-way head transplant situation (Dennett, 1978, p.313). It seems intuitively obvious that if your head is transplanted onto a different body *you* go with your head. The alternative is that you remain as the decapitated animal, so long as it is able to survive the process (by being hooked up to a life support system, for example).

Imagine the following scenario: at some point in the near future we develop the technology necessary to carry out head and nervous system transplants - the severing and reattaching of nerve endings is handled by surgeons wielding highly accurate computer-supported instruments. Such operations are routine: in fact, a number of your friends and relatives have undergone such a procedure as a form of life extension. Unfortunately, all of your major organs other than your brain have begun to fail owing to an illness of some kind. Not having a cure for this illness your best bet is the following: a routine operation can remove your head (and nervous system) from your body and attach it to a healthy one. Luckily for you, a body-donor recently met a gruesome end in an accident that resulted in the loss of his head. There is now, then, a perfectly healthy body being kept in an incubated state, awaiting a new head. Is it really reasonable to suppose that you might be reluctant to undergo the operation on the grounds that you might remain as a disembodied animal? Would your survival instincts really be aroused in such a way, given the many perfectly happy post-transplant people you have conversed with? We can suppose that for unknown reasons it has been discovered that the procedure is much more likely to be successful (physiologically speaking) if the patient is kept conscious throughout. The experience of being removed from the top of your body and placed onto another one would be, in more ways than one, unnerving. We can therefore suppose that copious drugs are given to you to induce a calm and tranquil state of consciousness. You are singing an old favourite tune when the surgeons begin the delicate task of removing your head. You are surprised by your reaction to being passed into the hands of a nurse: you find it amusing and mutter something about rugby. You are still in your drug-induced high when your reattachment begins and remain so throughout the whole process. Given that each (surreal) stretch of experience directly flowed into the next, is it not obvious that you have survived this operation? Is there any plausible sense in which, looking over at your ex-body post-operation, you might worry that you had not survived? It is difficult to take such a possibility seriously.

Animalists have their responses to this, of course, but the point is clear: it seems that a perfectly intuitive account of the self is a rare thing indeed. In defense of Animalism, Olson claims the following:

> Every interesting metaphysical claim has unwelcome consequences, and as unwelcome consequences go, these seem to me pretty mild. Couldn't we just be wrong about who would be who in imaginary brain-transplant cases, just as we can be wrong about who would be who in other bizarre cases (amnesia, brainwashing, duplication) where the usual patterns of evidence break down? (Olson, 2007, p.47)

We might well be wrong in such cases but the fact remains: our intuitions can be challenged deeply even when the theory on offer is as naturalistic as Animalism. Olson gestures at a further important point in claiming that there are unwelcome consequences for "every interesting metaphysical claim". This is a very general statement and it would be difficult to prove conclusively but it does seem a reasonable one when applied to the domain of personal identity. In our typical non-philosophical moods the conception we have of ourselves on a day to day basis is not the result of deep theorising: if we think about our essential nature at all it is not usually as part of an attempt to get at the necessary and sufficient conditions for personal persistence. Our everyday thoughts about ourselves are for the most part socially orientated: we think about ourselves in terms of our relationships to other people and in terms of our overall life narrative. That there would be a clash between this general conception of ourselves and a purposefully specific view of our most essential nature is surely not that surprising. If it is less palatable than the counterintuitive facts we learn and accept concerning the nature of the objects in the world around us, perhaps this is due in large part because it is about *ourselves*. Whatever the underlying reasons for this, it seems fair to say that discounting *The Ephemeral Self* purely on the grounds that it strikes us as a very strange view is not a justifiable move.

§5.11 Memories and intuition

The second issue to consider is the following: in certain ways *The Stream Thesis* is very much in accordance with our intuitions concerning personal persistence. It matches our intuitions concerning our moment-to-moment survival (and does so better than any competing view) and on this point *The Stream Thesis* is intuitive in a *phenomenologically* supported way: when we attend closely and carefully to our stream of experiences we notice that they flow into each other seamlessly. That co-conscious experiences such as this are consubjective is a highly plausible claim. In other words, the reason for our intuitive acceptance on this matter is a solid one. What of the reasons we have for finding *The Stream Thesis* unpalatable? As we saw earlier, there may be general reasons for the occurrence of psychological friction when we are confronted with a new picture of our nature. But, specifically, for what reason do we find it hard to accept that what we are now did not exist yesterday? In addition to common sense views about personhood a significant part of the answer is memory.

In recalling an experience we take it as evidence that we have existed prior to the present moment. If you remember having breakfast yesterday you do not just remember someone-or-other-having-breakfast: you remember *yourself*-having-breakfast, or so it seems. How could you recall you-having-breakfast if there was no "you" yesterday? A powerful impression of having existed yesterday is certainly had, but it is arguably not as strong or robust as the impression of continuity we feel when we attend carefully to the flow of our experiences. Unless we subscribe to some kind of direct realist account of memory (and hold that in recalling some event we are connected directly to the past) then we must accept that a memory of having breakfast yesterday is only a representation of that experience. We can now ask: in what way does this representation necessarily entail that the self currently recalling a previous experience is the selfsame one *represented* as having existed previously? The following claims from Bertrand Russell are relevant:

> It is not logically necessary to the existence of a memory-belief that the event remembered should have occurred, or even that the past should have existed at all. There is no logical impossibility in the hypothesis that the world sprang into being five minutes ago, exactly as it then was, with a population that "remembered" a wholly unreal past. There is no logically necessary connection

> between events at different times; therefore nothing that is happening now or will happen in the future can disprove the hypothesis that the world began five minutes ago. (Russell, 1995, p.132)

Russell's essential point stands even if we take it to be an analytical truth that I can only remember my previous experiences (i.e. experiences that *I* underwent): we need only change his "memory-belief" to an "apparent-memory-belief". Following Parfit (1984, p.220) we can label such apparent memories "quasi-memories" or "q-memories": they are qualitatively indistinguishable from genuine memories (and they are about previous events) but there is no conceptual requirement that the same subject is involved. On the account offered here, then, my memories of earlier today are genuine memories, but my memories of yesterday are only q-memories. There is, crucially, nothing in either the phenomenological content of the memory experience or the causal mechanisms responsible for it that *necessitates* the identity of the subject over the represented time frame: that we can easily conceive of the world having only sprung into being five minutes ago speaks to this point.

The psychological tension we encounter when confronted by *The Ephemeral Self* account's claims is not based on a commitment informed by philosophical analysis: the claim that memory necessitates an identity between the remembering self and the self remembered is not the cause of the trouble. It is based on our *interpretation* of our memories, which in turn is informed by the narrative we tell to ourselves about ourselves. This is not a particularly strong reason for holding an account to be damagingly counterintuitive. It should not be granted as much weight as our intuitive acceptance of *The Stream Thesis*' moment-to-moment persistence conditions: this acceptance is grounded in our phenomenology *itself*, as opposed to being grounded in an interpretative *belief* about what our phenomenology represents when it takes the form of a memory. We may find it hard to believe that we did not exist yesterday, but it is perfectly *conceivable*. It is not unbelievable in the way that becoming separated from our current stream of consciousness is: the lack of believability here is directly linked to its inconceivability, not to a pre-theoretical conception of the self.

In fact Strawson has argued that not *everyone* shares the intuition that the self enjoys a long-term existence. He conceives of a psychological division along the following lines:

> Endurantist individuals are those … who in the daily living of their subjecthood … intuitively figure the self as something that has long-term diachronic continuity, something that was there in the remoter past and will be there in the further future. Impermanentist individuals by contrast, are those … whose natural, regularly, lived sense of things is that the self that they now experience themselves to be is not something that was there in the remoter past or something that will be there in the further future, although they are of course fully aware of their long-term continuity as human beings considered as a whole. (Strawson, 2009, p.221)

Strawson hypothesises that this contrast in outlook may be the result of "a fundamental difference between human beings, with its roots in brain chemistry and organization, in genetically determined differences in individual temperament or general mental style" (Strawson, 2009, p.221). It would be interesting to see a large-scale questionnaire on this issue, to try and put a number on the percentage of people in the general population belonging to each camp. It would also be interesting to see just how recent a self the majority of Impermanentists would be naturally willing to disown: having the sense that the self of my childhood is distinct from my current one is surely more common than having the sense that the self of yesterday is gone. Perhaps such a prediction would be surprisingly refuted: Strawson, for one, claims to quite naturally conceive of himself as having not existed yesterday. He uses "I*" to refer to the self considered as an inner mental subject:

> If I engage in the philosophical exercise of trying to reach back to some sort of part of yesterday's consciousness … and manage to come up with something, I will certainly judge that it 'belongs with' today's consciousness in so far as it is consciousness on the part of the same single human being that I am … I know for one thing, that I can't reach back to anyone else's consciousness in that from-the-inside way. But I don't thereby feel that it belongs with my present consciousness in such a way that I think it was I* who was there yesterday … It feels remote. Nor do I judge, or feel, that it is I* who was there yesterday. On the contrary, I judge - feel - that I* certainly wasn't there. (Strawson, 2009, pp.225-6)

Strawson doubts that he is unusual in this respect. Whether he is right or not is something we will have to remain agnostic on for now. It seems, however, that we can say this much: some people naturally have the sense that the particular mental subject that they are did not exist in the remote past, regardless of what memories they may hold. If they were also persuaded by the arguments leading to *The Stream Thesis* it seems likely, therefore, that they would not find it entirely counterintuitive. What of Endurantist individuals?

I, myself, am one: I have the natural sense that the self I am today existed as far back as my memory can go. This is true even when I purposefully keep in mind the view of myself as an experiential being. As such, the conclusions of *The Stream Thesis* do not *naturally* strike me as intuitive. If I think through it carefully, keeping in mind the arguments in support of *The Ephemeral Self* account, it *can* seem plausible and believable. However, even after rehearsing such arguments, my natural disposition is not altered: if I recall a memory from childhood it still very much strikes me as being about *me*: the being I am now. This contrast between intellectual belief and what we might call "experiential belief" (a lived sense of something's being the case) is not, however, a particularly new, or damaging, phenomenon. Determinists are all too familiar with it: an intellectual acceptance of the lack of genuine alternatives does nothing to dispel the sense that on choosing to do *x* one could have instead done *y*. When we are seeking a believable account of personal identity these two different notions of belief are important. If an account delivers only intellectual belief this does not mean that we cannot recognise ourselves *at all* in it, only that it might be a difficult picture to assimilate into our daily lives. There is no good reason to think that this presents a serious problem for *The Ephemeral Self* account: many theories demand intellectual belief without changing our typical, day to day, ways of thinking much. Indirect realism is a good example: we may be, with good reason, convinced that we are not in direct contact with the external world, but this consideration rarely makes an appearance in our daily lives. This is simply to accept that in some ways the models we have of the world and ourselves are inaccurate. Given the natural evolution of our brains this is not a surprise: a correct appreciation of the fundamental facts concerning our identity is not something that would have given us a survival advantage (*especially* if it is true that you will not survive unconsciousness).

What of the response that *The Ephemeral Self* invites when we look forward in time? What does it mean to say that what you are will cease to exist the moment you become unconscious? There are reasons to think that this claim is less threatening, both to *The Stream Thesis* and to our own psychological wellbeing, than it might at first appear. This reconceptualisation of death is part of a wider set of implications brought on by *The Ephemeral Self* account. There is good reason to think that although we must give up a good deal in accepting the theory, there are also significant gains to be made. We now need to consider these.

6: The Personal Copernican Revolution

The Ephemeral Self account, if accepted, demands a significant change in how we view our own nature: *The Stream Thesis* changes how we view our nature considered over time and, in doing so, demands a reconceptualisation of death. Even before this, however, *The Exhaustive Constitution Thesis* challenges some prior conceptions we have of ourselves and, as will be shown, suggests that our basic self-experience is in fact very different from how we typically take it to be. When this latter point is directly and experientially recognised by a subject it can fittingly be described as a kind of personal Copernican revolution. In considering this we will complete our defence of *The Stream Thesis'* strange implications concerning personal death.

§6.1 Death and myself

Just how palatable is a theory that entails that what you are will cease to exist the moment you become unconscious? The first point to make is the following: the claim that "I will cease to exist tonight" can be made with different intentions, i.e. the entity referenced is not necessarily the same in all utterances: "'I' is not univocal. We move naturally between conceiving of ourselves primarily as a human being and primarily as some sort of persisting inner subject" (Strawson, 2003, p.286). This is an important point and worth considering in detail:

> I reject the assumption that 'I' is univocal in the thought or speech of any given individual. The reference of 'I' standardly shifts between two different things in my thought and speech and in the thought and speech of others. Sometimes 'I' is used with the intention to refer to a human being considered as a whole, sometimes it's used with the intention to refer to a self - two things that have quite different identity conditions … To say this is not to assume that selves exist. It's simply to report a fact about how the word 'I' is used. (Strawson, 2009, p.6)

We have in fact already seen how easily the referent of 'I' can shift: recall the "body-swapping" thought experiment of §5.1 (pp.113-4). In entertaining the scenario you were employing the "I" concept: you were considering the question "where will *I* be after the psychological change?" in an attempt to root out your most fundamental intuitions concerning your own personal survival. Entertaining the outlandish idea that your stream of consciousness might come to be sustained by the brain in the head of another body required conceptualising yourself as no more than an "inner persisting subject", as Strawson calls it. We saw that there was no difficulty in *conceiving* of such a scenario, regardless of its presumed nomological impossibility. Quite how often we think of ourselves as "inner subjects" is difficult to know, but that it can and does happen seems hard to disagree with. If we can think of ourselves in such a way this is because we can experience ourselves in such a way: we can have the experience of being an inner subject without explicitly considering ourselves as human beings. Strawson elaborates:

> I mean the experience that people have of themselves as being, specifically, a mental presence; a mental someone; a single mental something or other. Such Self-experience comes to every normal human being, in some form, in early childhood. The realization of the fact that one's thoughts are unobservable by others, the experience of the sense in which one is alone in one's head or mind, the mere awareness of oneself as thinking: these are among the very deepest facts about the character of human life … They are vivid forms of Self-experience that are perhaps most often salient when one is alone and thinking, although they can be equally strong in a room full of people. (Strawson, 2002, p.104)

The relevance of this to the issue at hand is obvious: the claim that you will cease to exist as soon as you lose consciousness can be taken to mean quite different things, depending on the kind of self-concept employed at the time of consideration. The believability of such a claim alters depending on how the claim is heard - there is, after all, all the difference in the world between hearing the following:

1. The particular stream of experiences that you are currently conscious of will end tonight.

Or,

2. The human being that you take yourself to be (with all of its memories, desires and social relationships) will end tonight.

If care is taken to bear this distinction in mind then the claims of *The Ephemeral Self* account appear far more reasonable. The problem, of course, is that these two meanings are easily conflated: it is quite hard to consider the claim that "what I am will not exist tomorrow" and remain focused on the use of "I" denoting your experiential nature only. One's mind is pulled in the direction of the social conception of oneself: the human being considered as a whole. Trying to picture the continued existence of this organism, whilst simultaneously accepting that *you* (this particular inner subject) will be absent, is no easy task. It seems reasonable to assume that this difficulty does not merely stem from normal habits of self-conceptualisation, but that the typically negative connotations of personal death are also playing their part. Interestingly, such connotations are not entirely merited if indeed *The Ephemeral Self* account of our nature is accurate. This means the following: if we can succeed in truly believing the claims that this account is putting forward, a significant factor in our finding it counterintuitive is lessened.

Consider the ways in which this account reconceptualises personal death from both the first-person and third-person perspectives. With regards to the latter: it is occurring much more often than we previously believed but in such a way that, ordinarily, *no one even notices*. Aside from positing universal eternal life, it is hard to think of a more significant way of diminishing the ominous character we attribute to our own personal demise. Every (non-vicariously)[34] conscious being existing on Earth last week has since died, and no one has noticed anything out of the ordinary: society has continued on in just the usual way. Of course, from this third-person perspective, the *biological* death of a human being is no less an event of finality and importance. The human being (the consciousness producing organism) has ceased to exist and, with that, so too have its social relationships. This is an important fact that it would be strange to ignore but the following point remains: the (genuinely) conscious being

[34] As defined in §3 (pp.78-81).

that has at that moment ceased to be (i.e. the self-intimating stream of experiences) had only been in existence for a relatively short time. Such a consideration will of course do nothing to console those in mourning over the death of a human being, but it speaks to the following: from the first-person perspective the significance of biological death is *fundamentally* altered if *The Ephemeral Self* account is accepted. Unless the organism currently producing your experiences dies today, its biological death is not something that is going to have any effect on *you*. You, as the stream of experiences occurring right now, will have long gone.

If this strikes you as swapping one piece of bad news for another then consider this: the conscious being that was produced by your C-system (or brain) yesterday, who you are currently prone to identify yourself with, has ceased to exist and, crucially, this fact is of very little importance to you *now*. Similarly, the end of you (of this particular stream of experiences) will be of no significance tomorrow for the conscious being that shares your C-system: it will *seem* as if the self-same entity has survived the night. Is this state of affairs a depressing prospect? If we are prone to thinking of ourselves as fairly long-lived beings - as an entity that has experiences, but who stands apart from them - then it does imply a great loss. As we shall see, a consideration of the phenomenological situation entailed by *The Exhaustive Constitution Thesis* is relevant here: a true appreciation of its claims concerning self-experience can make *The Ephemeral Self*'s reconceptualisation of death much more acceptable on a personal level.

In a sense *The Ephemeral Self* is an error theory: it claims that, although we typically believe ourselves to be more than just our currently occurring stream of experiences, we are mistaken. There is a further aspect to this: in becoming aware of the error, and in purposefully attending to some phenomenological facts relating to it, there is good reason to think that a certain degree of general psychological suffering can be alleviated. On both of these points *The Ephemeral Self* shares a good deal with the claims and goals of much of Indian Philosophy, even though the account presented here refrains from some of the bolder metaphysical claims that can be found in the latter.

§6.2 A middle way

Tentatively, as we enter into somewhat more speculative territory now, *The Ephemeral Self* can be used to present a naturalistic soteriology (whilst also giving due credit to our most fundamental intuitions concerning our own moment-to-moment persistence). On its soteriological features, it sits well within the following wider context:

> Irrespective of the often considerable differences between their metaphysical doctrines, many of the major philosophical schools of India agree in their basic assumption that, in order to become aware of one's own true nature, one has to inhibit one's self-consciousness in the usual sense, namely one's 'ego-sense' (ahaṃkāra, literally 'I-maker'). The normal way we are aware of ourselves - that is, our self-awareness as a distinct psychophysical entity with particular characteristics and abilities, formed by a personal history, standing in manifold relations to other things and persons, etc. - is in this view really the construction of a pseudo-self that obscures what we really are … for Buddhism this means that the spiritual aim is to realize that it is an illusion that something like a self exists at all, for 'orthodox' schools such as Advaita Vedānta or Samkhya and Yoga, liberation lies, on the contrary, in becoming aware of the true self (ātman or puruṣa). (Fasching, 2010, p.193)

Fasching goes on to present an enlightening interpretation of the phenomenological situation confronting a conscious being engaged in the above practise. A significant portion of Fasching's claims is both philosophically sound and amenable to the account being put forward here. There is, however, room for disagreement. Working through and tackling his various arguments will shed light on *The Ephemeral Self* account and go some way to addressing its counterintuitive treatment of personal death. Further to this, by placing it somewhere between the metaphysics of the orthodox schools and their Buddhist rivals a good case can be made for the following: *The Ephemeral Self* is, in actual fact, a relatively *moderate* theory in this context.

Fasching's stated goal hints at our desired middle-ground:

> I would like to cast, from a phenomenological point of view, some reflections on what this overcoming of the ego-sense strived for by these traditions could possibly mean, and will try to vindicate the view of Advaita Vedānta that it does not amount to a dissolution of oneself into a mere flux of substrate-less transient phenomena, but rather to a realization of one's self as something that changelessly underlies this flux. (Fasching, 2010, pp.193-4)

Broadly speaking, and rephrasing to suit the context of our discussion, the orthodox schools of thought (such as Advaita Vedānta) hold that the self persists throughout the entirety of a human life (and, indeed, thereafter). The various forms of Buddhism, on the other hand, hold that there is no such thing as genuine identity over time: there is no self at all, i.e. the no-self theory (anātman) is true. *The Ephemeral Self* occupies a middle-ground: there is no eternal (or even biological-life-long) self, but there *is* genuine personal identity over time (although for a shorter duration than we typically believe). Investigating Fasching's attempted vindication of Advaita Vedānta will help to both elucidate this middle-ground and show why it is the strongest position to hold. In a nutshell, we will see good reasons to reject Fasching's bracketed claim below whilst agreeing that there is something importantly true, though perhaps somewhat difficult to get at, concerning his interpretation of "witnessing":

> the claim against the Buddhists is not that there has to be some entity in addition to, and behind or beyond, our experiential life as its substrate, but that there is a stable element within it - yet not as some invariant content or content-constellation we could experience (such a thing is indeed not to be found), but as the very process of experiencing itself, as the permanence of 'witnessing', in which everything we experience has its being-experienced, and which is the constant ground of our own being. (Fasching, 2010, p.194)

To the first point we have already seen that there *is* indeed an "invariant content" to be found within a typical stream of experiences: co-consciousness. Co-consciousness is an *experienced* relation between experiences. Fasching, on the unity of consciousness, claims that "nothing on the content-side can do this job" (Fasching, 2009, p.142). In §4 (pp.82-108), we saw strong arguments to the contrary that we won't rehearse here. An important motivation for Fasching's reluctance to accept the

role of phenomenal relations here is the following: if one accepts that consciousness can unify itself *by itself*, as it were, then the theoretical need for a transcendental subject vanishes. The problem with this, for Fasching, is that the following thought-experiment induced implication is unacceptable:

> It appears to be perfectly conceivable that this very experience with all its relations to other experiences of the same stream of consciousness, to this body and to the rest of the world, could have existed without the 'I' which experiences it being *me*. This seems to be a contingent … fact. (Fasching, 2010, p.199, original italics)

This (apparent) contingency is, for Fasching, entailed by the following:

> This 'who' of experiencing is an additional fact with regard to the experience and its phenomenal character: No facts whatsoever about an experience or its 'what-it-feels-like-ness' can ever imply its being experienced *by me* (except, precisely, that it is *I* who experiences it). (Fasching, 2010, p.199, original italics)

The problem with this claim is that it is only if I hold myself to be *distinct* from my currently occurring experiences that I can coherently imagine undergoing a different set of experiences (with none in common with my current stream) *whilst remaining myself*. The impression that I am distinct from my current stream of experiences is precisely what is being shown to be inaccurate by *The Exhaustive Constitution Thesis*. Bayne has the following to say on the issue:

> Prima facie, it seems to be coherent to suppose that one could have enjoyed a stream of consciousness that had no experience in common with those that one has actually enjoyed … The naïve phenomenalist may need to accept that it is incoherent to suppose that one could have had radically different experiences. This may be an objection to naïve phenomenalism but it is not a knockout blow, for a good case can be made for thinking that *any* plausible account of the self will need to reject some of the modal intuitions that surround the self. (Bayne, 2010, p.283, original italics)

As we saw earlier, our intuitions concerning selfhood seem to be challenged whenever we attempt to specify our persistence conditions in a precise way. The arguments in support of *The Stream Thesis* are, on their own, good reason enough to hold that Fasching's above thought experiment is incoherent. Briefly put: at the present time I am exhaustively constituted by the currently occurring experience (*x*) of which I am (non-inferentially) conscious, owing to the fact that *only x*, as self-conscious and self-intimating, can be conscious of *x* in such an immediate fashion. Owing to the phenomenal unity that exists between *x* and the rest of the experiences in its stream, I am identical with the stream of consciousness of which *x* is a part. As *this* stream of experiences I could not have been an entirely different stream of experiences: it simply makes no sense to say so. This stream of experiences is *this* stream of experiences. If you can seemingly imagine yourself enjoying a stream of experiences sharing *zero* content with your current one, then you are not imagining *yourself* undergoing change: you are picturing yourself as the *bearer* of this stream - as something distinct from it. This is not to deny, of course, that we can coherently imagine having woken up today and decided to spend the rest of the day in bed: if we rewind our stream of experiences in our imagination and picture a different set flowing on from our earlier ones, we have imagined a possible world in which we have different experiences. The crucial point is that such experiences would have been directly connected to one or more of the experiences of *this* stream occurring *now*. This is different from supposing that "I" could have undergone the experiences of, say, the President of the United States as they were yesterday: this is simply incoherent. It would be even better however, if we can show where Fasching goes astray in his own terms. For this, we need to consider the widely defended notion of "mineness".

§6.3 Mineness

This aspect of experience is held by Fasching to, in part, do the work of individuating different subjects of experience (Fasching, 2008, p.133). We need to investigate, then, whether or not "mineness" can indeed perform this function and, furthermore, if the necessary work can be done without it. In doing so, it will become clear which aspects

of Fasching's account of "becoming aware of one's true nature" (Fasching, 2010, p.193) are favourable and which we should attempt to do without. Fasching makes his case as follows:

> In my experiencing them, my experiences are given to *me* in a totally different way than to anyone else, and are in this sense the experiences *I* have or *I* experience … They are experienced by me and not by you … The "I" is not to be found as some observable object, but in the fact of the mineness of my experiences. (Fasching, 2009, pp.132-3, original italics)

Dan Zahavi (2005b) employs the same notion for similar purposes. On the shared features of different experiences he says the following:

> One commonality is the quality of *mineness*, the fact that the experiences are characterized by first-personal givenness. That is, the experience is given (at least tacitly) as *my* experience, as an experience *I* am undergoing or living through. (Zahavi, 2005b, p.16, original italics)

As an act of descriptive (i.e. not transcendental) Phenomenology, Zahavi's claims seem reasonable enough (depending on how we interpret them). A problem develops when, instead of "mineness" being posited as a phenomenological feature of experience, it is held to be the mechanism responsible for individuating subjects:

> Whether a certain experience is experienced as mine or not, does not, however, depend upon something apart from the experience, but exactly upon the givenness of the experience. If the experience is given originarily, in a first-personal mode of presentation, it is experienced as *my* experience, otherwise not … the particular primary presence of the experience makes it mine, and distinguishes it from whatever experiences others might have. (Zahavi, 2000, pp.60-1, original italics)

It seems that "mineness" is here employed to mean two very different things: firstly it refers to the *way* an experience is made manifest for its subject. But it also appears to be referring to what it is that makes an experience belong to subject *A* as opposed to

subject *B*. In other words, it is a structural feature of my experience *and* it is what makes my experience *mine* as opposed to yours. A serious problem looms. Fasching, referencing David H. Lund (2005, p.118), is aware of it:

> The question of what it means to say that it is me who has this present experience cannot be answered by referring to its property of "mineness" (i.e. first-personal givenness), for *every* experience that ever has been experienced possesses this feature. But only those experiences that are first-personally experienced by *me* are mine. The mineness of my experience is not *some* mineness but *my* mineness. What is this mineness of mineness? (Fasching, 2009, p.138, original italics)

When "mineness" is employed in such a way its defenders are stuck on the horn of a dilemma: either the notion can no longer individuate subjects (owing to the fact that it is a quality inherent in all experiences) or it has to be combined with a further feature that is itself little more than a stipulation. Zahavi opts for the latter option: on his account an experience is mine because it is characterised by a "mineness" that is instantiated in a particular primary presence: *my* primary presence.[35] But what exactly *is* this primary presence? And what makes it *my* primary presence? It seems to be no more than a theoretical placeholder for the task that "mineness" cannot fulfill on its own.

The Ephemeral Self account, employing only the phenomenologically sound notion of co-consciousness, deals with the issue much more efficiently. Consider your current experience. Remember to keep in mind that although language demands that it is most naturally described as "your" experience it is in fact not something distinct from you: during its occurrence you are exhaustively constituted by it. In saying that this experience is "given to you" in a way that it is not "given to me" Zahavi and Fasching are describing the following fact: my experience is not something you are conscious of in the way that you are conscious of your own. You may infer that I am having an experience by observing my behavior (among other things) but you do not infer your own: it is immediately present. The same is true from my perspective and

[35] Fasching also combines "mineness" with a further feature: the "dimension of first-personal presentation which has no other essential property than being me" (Fasching, 2009, p.146). This seems, ultimately, to be the same concept.

this much we should certainly agree with. This state of affairs, however, can be described quite adequately without recourse to "mineness" in either of its uses. Call the experience of which you are (non-inferentially) conscious "*x*" and the experience of which I am (non-inferentially) conscious "*y*". The state of affairs is simply this: *x* and *y* are not co-conscious. They are self-intimating and self-conscious, and they do not enter into the immediate phenomenal sphere of each other.

The Stream Thesis, with its notion of co-consciousness, accounts for the claim that my current experiences are "given to me" in a way that other currently occurring experiences are not, without motivating the notion that there is something experientially *unique* about the subject of this particular stream. There is *this* stream and there is *that* stream, and that is all that needs to be said: there is no special feature of my experience (unique to "me") that is responsible for individuating streams. We can put this point in the following way: upon having specified all of the phenomenal relations between every experience in existence at this very moment, we have *already* individuated every stream of consciousness from every other one. If such a scenario leaves us with the impression that something is left out of this picture then it is because we are still thinking of ourselves as *distinct* from our conscious experiences, as an entity that stands apart from its phenomenology.

The point can also be put the following way: suppose that, for the sake of argument, each stream of co-conscious experiences does indeed have a "primary presence" unique to its subject. Let's further suppose that, as Zahavi and Fasching both agree, each experience is self-conscious and self-intimating in the way identified in §1 and §2 respectively. You are acquainted with only those experiences that are given through your particular primary presence, and the same holds for me. Now let's suppose that this primary presence suddenly ceases to exist but that the relevant experience producing mechanisms continue to produce experiences with the same co-conscious relations: immediately, and without warning, experiences are simply given "anonymously", which is to say they do not belong to any particular primary presence but instead just appear for their respective subject. Would we notice the difference? The experiences that were previously given through your "primary presence" are still very much *not* co-conscious with the experiences previously given through my "primary presence". In other words, what can the notion of "primary presence" account for that streams of co-conscious experience cannot? If it is merely the intuition that "I" could have had a different set of experiences (with none in common

with my current stream), then this is a weak motivation. We should interpret the following intuition differently from how we may at first be inclined to:

> Strangely, it appears to be a *contingent* fact that one is the person one is … No third-personally formulated feature of the person I am (be it his body, personality, or experiences) seems to logically imply that this person is not just "an I", but me… (Fasching, 2009, p.141, original italics)

Instead of taking this to mean that there must be more to being me than the occurrence of this particular stream of experiences, we should accept that indeed there is never anything more than, as Fasching puts it, *an* I. The sense that there is something experientially unique about this "I" is simply a mistake. We can either accept that we are misled as to our nature in this respect or else become committed to the mysterious notion of "primary presence" which accounts for nothing (aside from a questionable intuition) that co-consciousness cannot. Furthermore, such a notion has the suspicious property of exclusively solving the problem at hand without itself being particularly scrutable:

> No features of my experiences - no matter whether intrinsic or relational - can imply their being mine except their having their manifestation in this very dimension of first-personal presentation which has no other essential property than being me. (Fasching, 2009, p.146)

In a similar fashion to the issue of temporal consciousness, if we can focus on the phenomenological data and avoid positing theoretical constructs such as this then our theory is stronger for it. It is worth bearing in mind just how intuitive *The Stream Thesis* is, prior to its dealings with death:

> The view that the oneness of "one mind" means that various experiences are unified, and that this unification is due to some inter-experiential relations is certainly not utterly implausible. On the contrary, as a matter of fact, prima facie, it seems to be *nearly self-evident* - how could it be otherwise? Yet on closer inspection, as the debate about personal identity brought to light, it turns out to have highly counter-intuitive implications. (Fasching, 2009, p.135, my

italics)

As we saw, however, a close inspection should include specifying the phenomenal relations between experiences as opposed to exclusively dealing with physical and psychological continuity. This thought goes some way to addressing another of Fasching's problems with views of the reductive kind:

> These implications are due to the fact that concepts like "connectedness" or "continuity" allow for degrees and also, at least in principle, for splitting and merging … Experiences can be more or less unified, borderline cases are possible where only an arbitrary decision can determine whether something should or should not count as belonging to a unit of bound-together items. But can experiences be more or less mine? Can there be borderline cases of mineness? Is it not necessarily the case that each experience is either experienced by me or not, without any in-between? And is not the question of whether it is the one or the other a real difference and not just a question of referring to the same matter of fact in different ways? (Fasching, 2009, p.135)

The Stream Thesis is compatible with Fasching's claim that experiences cannot be more or less "mine": an experience is either co-conscious (directly or indirectly) with my current one or it is not (i.e. it is either "mine" or it is not). As we saw, this is one advantage that phenomenal continuity has over psychological connectedness: it does not result in indeterminacy.

It is true that the imaginary scenario in which a stream splits into two (or more) presents a difficulty for an experiential account. The same issue can, however, be turned back onto Fasching: imagine that your current stream of experiences suddenly splits into two. Suppose that branch *A* retains your "primary presence" and branch *B* does not. According to Fasching and Zahavi experience of our kind is necessarily given through a primary presence, so perhaps they would say that the experiences in branch *B* are now given through a *new* primary presence? Consider this scenario from the perspective of the subject moving from the original stream into the *B* section. If consciousness was retained throughout the process then it seems that the subject of the experiences in branch *B* was also conscious during the original phase of the stream: any given experience in the *B* phase is diachronically co-conscious with any

given experience in the original phase, owing to the overlapping specious presents that exist between them. According to the current interpretation of this state of affairs, however, a new primary presence (and hence a new *subject*) has begun. If primary presences do the job of individuating subjects then there is a dilemma: either a subject *can* come apart from its stream of consciousness (this is what happens to the subject of the original phase of the stream that ends in *B*), in which case such an account goes against one of our most strongly held intuitions concerning personal persistence, or else the subjects in phases *A* and *B* share the *same* primary presence and are therefore in fact identical despite enjoying different branches of the stream. In facing the latter option a primary-presence account is in the same boat as a reductionist account of the experiential kind.

Perhaps an overlapping of some kind could solve the problem: Fasching might hold that the experiences in the original phase of the stream were in fact given to two different subjects (each with their own primary presence) who then separated upon the division of the stream. How much sense this makes is unclear, but an overlapping move of this kind can also be made by an account that only makes reference to experiential-relations. In short, unusual scenarios such as the splitting of conscious streams present a problem for *both* sides of the debate. This is not particularly surprising given that our basic intuitions concerning personal identity are rooted in *typical* experiences: stretching our everyday concepts by considering unusual scenarios is predictably going to provoke friction.

§6.4 The luminosity of consciousness

There are a number of reasons, then, to refrain from following Fasching in his defense of "mineness" and "primary presence". As we will see, however, he *does* present a valuable account of the phenomenology at play for someone recognising his or her essential nature as consciousness. To a certain extent this is employed in an attempt to strengthen his anti-reductionist position but a good deal of his observations are convincing when considered in isolation from that project. As such, they can be used to address the issue of *The Stream Thesis*' counterintuitive treatment of personal death.

It strikes me that, in trying to grapple with the following issues, we are taken

somewhere near the limits of discursive thought. But it also seems that there is something of genuine worth in the following discussion and that we should therefore at least *try* to get to the bottom of it. With this warning in mind, we need to begin with Fasching's notion of "presence": he claims that "[what] Advaita Vedānta soteriologically aims at as the realization of the 'self' is nothing other than becoming aware of experiential presence (consciousness) as such" (Fasching, 2010, p.207). This occurs, for Fasching, through a process of "de-superimposition" whereby:

> Instead of identifying certain configurations of experienced contents as being 'oneself', one begins to experience oneself as the abiding experiencing itself (the taking place of presence) of any contents. De-superimposition means radically distinguishing oneself from all objects by no longer delimiting oneself (as an 'inside') as opposed to the objects 'out there'. One stops considering anything as being 'oneself' or 'one's own' … In the 'de-identified' mode of experiencing that is strived for, one completely lets go of 'oneself' and becomes nothing but 'seeing', without any distinct 'seer' standing apart from the 'seen' … One becomes aware of oneself precisely when one ceases to find oneself anywhere. (Fasching, 2010, pp.211-12)

Within this dense chain of claims there is much to both agree and disagree with. We need to dig down into the key concept of "presence":

> for the Advaitins - although they hold that mental states are manifest *essentially*, and not by virtue of being the object of some further, higher-order mental states - it is not adequate to say that they are immediately *self*-aware. Rather, they exist in manifesting themselves *in* the medium of the luminosity of consciousness, which is immediately self-revealed … while these experiences are permanently fleeting, conscious presence as such abides. (Fasching, 2010, p.202, original italics)

Fasching is aware of the problems that can come from drawing a distinction between individual experiences and consciousness; we are "obviously hypostasizing consciousness into a 'something' in addition to experience" (Fasching, 2010, 202). He claims, however, that there *is* a justifiable way to make such a distinction, "in which a

multitude of experiences can be the taking place of the same consciousness" (Fasching, 2010, 202). We have already seen the reasons Fasching gives for taking this route and found them wanting. As such it is difficult to agree with his view that "presence" is experience-unifying or that it is responsible for the individuation of subjects. We have also seen solid reasons to refrain from drawing a distinction between consciousness and its contents in §2 (pp.40-77) (the discussion in §2.8 (pp.70-77) is particularly relevant). There are, however, good reasons to think that his notion of "presence" when considered just in its mode of "luminosity" is getting at something extremely important (although elusive) with regards to the process of de-identification under discussion. If we can understand it without "hypostasizing consciousness" into something over and above individual (albeit connected) experiences then we may be able to use much of Fasching's interpretation to shed light on *The Ephemeral Self*'s soteriological possibilities, and thereby reduce the intuitive unease we might feel when considering its reconceptualisation of personal death.

We have in fact already touched upon the "luminosity" of consciousness that Fasching mentions. In agreeing with Dainton's *Simple Conception* of experience and holding that experiences do not require a distinct awareness in order to be conscious, we identified them as "self-intimating" or "self-revealing" or "self-*illuminating*". Putting to the side the reflexive aspect of these concepts for now, we need to consider the following: what does it mean for some phenomenal content to be "intimated", "revealed" or "illuminated"? The simplest though perhaps least informative way of providing an answer is the following: phenomenal content is luminous in the sense that it is *here*. It is manifested. It is immediately present for its subject in such a way that its existence is not open to skeptical doubt: the luminosity of an experience is its undeniable *appearing*. As such it is the shared property of all phenomenal content: it is what makes it *phenomenal* content in the first place. Revisiting the comments of Moore from our investigation into transparency, it is this very feature of conscious experience that he himself was grasping at:

> The term 'blue' is easy enough to distinguish, but the other element which I have called 'consciousness' - that which sensation of blue has in common with sensation of green - is extremely difficult to fix. That many people fail to distinguish it at all is sufficiently shown by the fact that there are materialists.

> And, in general, that which makes the sensation of blue a mental fact seems to escape us; it seems, if I may use a metaphor, to be transparent - we look through it and see nothing but the blue; we may be convinced that there is *something*, but *what* it is no philosopher, I think, has yet clearly recognised. (Moore, 1903, p.446, original italics)

Accepting the difficulty we face in trying to attend to (or conceptualise) such a feature, we saw that Moore went on to say the following:

> the moment we try to fix our attention upon consciousness and to see *what*, distinctly, it is, it seems to vanish: it seems as if we had before us a mere emptiness. When we try to introspect the sensation of blue, all we can see is the blue: the other element is as if it were diaphanous. Yet it *can* be distinguished if we look attentively enough, and if we know that there is something to look for. My main object in this paragraph has been to try to make the reader *see* it: but I fear I shall have succeeded very ill. (Moore, 1903, p.450, original italics)

I think that Fasching presents a powerful interpretation as to why we find it difficult to identify this feature of conscious experience. Building on the metaphor of Moore, Fasching phrases it as follows: the luminosity of experience "is not one of the 'seen' things but the 'seeing' itself" (Fasching, 2010, p.208). Such phrasing can seem to lead to a reification of consciousness, or the positing of a distinct awareness gazing down at phenomenal content. We need not, however, go down this route. It is, after all, only a metaphor and as such can be put to work in various ways.

It will serve us well if we can apply this seemingly abstract notion of "seeing" to our own presently occurring phenomenology. Consider your current visual experience of this page. If you were to be asked to catalogue what is appearing to you right now in this visual experience you could list the external content: the words on the page and the objects surrounding this document. You could list the various properties of these objects: their shapes, colors and relative positions. As we saw in the discussion concerning the ubiquity of inner awareness in §1 (pp.16-39) there is something else you must include: this experience *itself*. You are aware of the visual objects you are seeing and you are aware of your experience of them even though this is not itself something you *see*. Your phenomenology is simultaneously world-presenting *and*

self-presenting (in the sense that it presents itself). Now consider your auditory experience. The same applies: you are conscious of both the things represented by your phenomenology *and* your phenomenology itself. In short, as Strawson phrases it, "all awareness involves awareness of that very awareness" (Strawson, 2013, p.5), i.e. your conscious experience is self-luminous and "we can say that awareness takes itself as part of its own *content*; we can allow that this is one acceptable way to express what its self-intimation or 'phosphorescence' consists in" (Strawson, 2013, p.26, original italics).

Staying only at the phenomenological level of inquiry, for any two experiences we can either focus on what makes each of them the *particular* experience that it is or we can focus on the feature that makes each of them *an experience* at all: the property of luminosity. Crucially, this is not to say that content and luminosity are in fact separable: as we saw with Dainton's *Simple Conception* (§2.8 (pp.70-77)) there are very strong reasons to hold that these two aspects of experience are not, in reality, distinct. But luminosity alone *can* be teased out in thought (and made the target of explicit attention) owing to the fact that it is common to all experiences.

Having done our best to differentiate luminosity from the other properties of experience, we must now tackle the following: in what way is luminosity *itself* phenomenologically present? This is a particularly difficult question to answer for the following reason: the luminosity of a particular phenomenal content *just is* that content's property of *being phenomenologically present*. Luminosity, it seems, is not some particular phenomenal content that can appear but the very "*phenomenality*" of such appearances, as Fasching puts it (Fasching, 2008, p.467, original italics). Yet we clearly are, somehow, aware of it: it is what we are grasping at when we recognise that there is a common property shared by all experience, one that is elusive but undeniable.

Dainton offers a tentative treatment of this issue at the very end of *Stream of Consciousness* (2006). Having examined the nature of co-consciousness in a deep and detailed fashion, he speculates that this very relation might be the common property shared by all experience and claims that "there is no denying that co-conscious unity is *a* if not *the* distinguishing characteristic of the phenomenal" (Dainton, 2006, p.239, original italics). We have seen a good deal of arguments in support of the first part of this claim. That co-conscious unity might be *the* distinction between phenomenal and non-phenomenal properties is harder to agree with. Dainton suggests the following:

> That this mode of unity is a crucial feature of experience as we know it is easily appreciated: remove this unity from a typical stream of consciousness and what would be left? At most a myriad of instances of point-like quality, each so entirely isolated from the rest that, from the point of view of experience at least, they could as well exist in different universes. So might it not be that *any* intrinsic qualities unified by co-consciousness should properly be regarded as phenomenal in nature? (Dainton, 2006, p.239, original italics)

The difficulty with thinking through this admittedly speculative claim is that the scenario that we are being asked to imagine is unlike anything we have ever experienced ourselves: a stream of consciousness minus its unity. As such it is difficult to imagine "what would be left" in such a case. In one sense, this could be seen to support Dainton's tentative claim: perhaps we cannot clearly grasp what non-unified phenomenal qualities would be like because they would not *be like* anything at all experientially speaking i.e. they would not be *phenomenal* properties and we, in some sense, recognise this implicitly. Given that we are able to imagine all sorts of strange experiences and have no problem at all in so doing, even when such postulated experiences defy the laws of known physics (such as when we imagine traveling backwards through time), perhaps this is the very reason why we find it difficult to envisage non-unified qualities as phenomenal: they aren't.

We should be careful not to move too fast however, and - in fairness - Dainton does not. Perhaps such strange point-like phenomenal properties *are* possible, and we are unable to conceive of them simply because we do not have the conceptual-imaginative capabilities to do so. Consider the following: if we successfully imagine the removal of all unity from a typical stream of consciousness, it seems that we have also removed *time* from the picture. If, in trying to envisage an instance of point-like quality, you have in fact imagined a very brief pulse of pinpoint greenness, for example, then you have not removed all traces of unity from the scenario: the earlier section of this briefest of qualities is unified with the later section. To *successfully* imagine the absence of unity we would need to imagine a literally instantaneous quality. Arguably, however, an instantaneous *anything* is not something we can clearly conceive of. It does not follow, then, that such an entity cannot exist. It might show, however, that such an entity cannot exist without itself having some form of

temporality. Considered in this respect Dainton's claim is intuitive: experience seems to be the kind of thing that requires at least some degree of temporal extension in order to exist. If *an* experience is extended, then it is necessarily unified through time. This, however, only speaks to the first part of Dainton's claim: that co-conscious unity is *a* distinguishing feature of the phenomenal: it does not show that it is the *only* one. Luminosity might also be such a feature. Perhaps co-conscious unity and luminosity are in fact interdependent: in order to be unified by co-consciousness *x* needs to be luminous, i.e. *x* needs to *appear*, have luminosity, in the way that experiences like ours appear, and in order to appear in such a way *x* needs to have some form of temporality and thus will also necessarily be unified by co-consciousness. Dainton's claim does have the following advantage: we can articulate what we mean by "co-consciousness" in a more satisfying fashion than we can with "luminosity". This is not enough in itself, however, to settle the matter.

On the difficulty of thinking about and describing "the glow of experience" (Strawson, 2013, p.28) in the language of structure and relation, Strawson claims that "we need to acknowledge the inadequacy, and accept that it lies in the nature of discursive thought" (Strawson, 2013, p.28). A little light might be shed on this issue if we switch to the terminology of Zahavi and Fasching and say that content is "given" to its subject when it is phenomenally (and immediately) present for its subject. The luminosity of consciousness now becomes the property of "being given". In our usual dealings with the world, and even in our specialised thinking about philosophical problems, we are confronted with various kinds of things that are given - ideas, objects, and experiences - but we do not usually spend time dealing with the question of what it is for something to *be given* in the first place. Ironically, that things can be given at all is taken as given. It is perhaps not surprising then that when we do try and attend to the matter our usual ways of thinking seem to fail us. For Fasching, this is a habit we cannot easily break even when directly engaging in the philosophical problem of consciousness:

> To say subjective experience is actually particular neural events, given "introspectively", i.e. as we *subjectively experience* them, begs the question. The taking place of phenomenal givenness - which is, in my view, the core explanandum when it comes to consciousness - is simply presupposed. (Fasching, 2008, p.468, original italics)

We have strayed into the problem of the mind-matter relationship and will not now enter any further into it. It does seem to me, however, that Fasching has identified something important here. Bearing in mind the difficulty that language seems to have in reaching into this sphere, consider the following:

> So, by consciousness, we mean here the event of phenomenal presence of whatever is present. The distinction between "real" objects and those that only exist "in our mind" is a subsequent one in comparison. So in a way there are "subjective" and "objective" phenomena, but consciousness is *not* a subjective phenomenon, it is not an "inner world": It is the being-there of whatever kind of phenomena - whether "subjective" or "objective". Consequently, consciousness is not a phenomenon among phenomena but the taking place of the *phenomenality* of phenomena. (Fasching, 2008, p.467, original italics)

Fasching's claims are powerful and, I think, perfectly compatible with the *Simple Conception* of experience if we take them to be describing the luminous nature of *experiences*: luminosity is just the "being-there" of our self-intimated experiences. It is not a "medium" in which contents flow through, as Fasching has it (Fasching, 2010, p.22), belonging to a consciousness *distinct* from its experiences (Fasching, 2010, p.202) but an intrinsic *aspect* of any given experience itself: its self-intimating "glow". It is, however, not a *typical* phenomenal datum because it is not itself a particular content: it is the *appearing of* phenomenal content. With this in mind we can begin to appropriate Fasching's phenomenological claims into *The Ephemeral Self*'s soteriology.

§6.5 The ego

If you are identical with your current stream of experiences then your very own nature is comprised of these two (inseparable) aspects: qualitative content and luminosity. Furthermore, if the soteriological project under discussion here consists in a conscious process of de-superimposition, whereby a subject becomes increasingly (and explicitly) conscious of the luminosity of experience, then we need an account of

what it is like to apprehend this aspect of our own nature (and we also need to look at why it might be a *soteriological* project). We need to try to make sense of what it is, phenomenologically speaking, to become explicitly aware of this feature of ourselves. Although there are good reasons to disagree with Fasching's account of "mineness" and the individuating of streams (or selves) in terms of it, he does provide a good deal of phenomenological insight into the process of "de-superimposition" that we should agree with. In line with the views of Advaita, he holds consciousness to be distinct from its individual experiences. As such, he takes the process of de-superimposition to be a recognition that one is, as consciousness, not one of the various things that are appearing. Phenomenologically speaking, one recognises oneself as the presence in which the content is arising: one is none of the things witnessed (be they external objects or subjective experiences) but the "witnessing" of them all (Fasching, 2010, p.12). The task is this: can such a phenomenological account be justified (and made sense of) when consciousness is *not* taken to be distinct from individual experiences? I think that this might well be possible. A clue is to be found here:

> From the very beginning, consciousness is self-luminosity, and there is nothing to add to this. So meditation is not about looking at some hidden place and discovering something special there. Rather, the meditative process consists in unconstructing the usual ontifying self-apprehension: One becomes consciousness through and through, without attributing consciousness to an "ego". (Fasching, 2008, p.477)

Instead of interpreting the above process of de-superimposition (or de-identification) as a way to "become consciousness", we can propose the following picture: one becomes explicitly aware of the luminous *aspect* of one's experiences, and this recognition facilitates de-identification with the "ego". It is not that one recognises that one is distinct from one's particular experiences, but that one recognises an aspect of one's own nature that is not itself a *typical* kind of phenomenal content. In recognising this, it can become clear that one is not an entity that "has" experiences but that this "ego" is just *part* of the content of any given experience: an experience that, as well as being comprised of such content, also has the aspect of luminosity. Fasching's fundamental point is right: you cannot be aware of luminosity without *being* it - you are consciousness through and through - but this does not mean that

your nature is *exclusively* comprised of it: as the experience considered in its entirety you are also constituted by its qualitative aspect.

This is the personal Copernican revolution: one recognises oneself as the conscious experience through which an ego entity appears, as opposed to an ego entity that "has" experiences. My claim here is this: focusing one's attention on the aspect of luminosity can foster such an insight, without the need to identify oneself as a consciousness in any way *distinct* from its experience (and without positing a unique feature of "my" consciousness that individuates it from "yours", as Fasching does). One recognises that one cannot be identical with the ego-content for whom objects are seemingly presented, given that the luminous experience as a whole (of which the ego-content is only a part) is something that *you* are aware of. This is an important point and worth looking at more closely.

The term "ego" can, and has, been used to mean various different things. In the context of the present discussion the ego is the entity that we naturally, and incorrectly, take ourselves to be. This, at present, is not a particularly useful definition and therefore needs elucidating. Before doing so, it is worth bearing the following in mind: in speaking of the "entity" that we naturally take ourselves to be we may not be referencing just one thing. As we have seen, there are different ways that we can conceive of ourselves and, although we may use one pronoun to tie all such conceptions together, such phenomena may be by their very nature varied and, perhaps, to a certain extent vague. As Hume himself noted: "in common life 'tis evident these ideas of self and person are never very fix'd nor determinate" (Hume, 2007, p.127). Having said as much, we can follow Miri Albahari (2010) in her articulation of the specific *experiential* features of this ego "entity", and then consider its other aspects in relation to this. Albahari begins by giving a general description of the self-conception at issue:

> In essence, then [sic] this commonly assumed self is a unified, unbrokenly persisting subject of experience with personalized boundaries and a perspective on the world. It is a thinker, owner, and agent that stands behind, and is somewhat in charge of, the stream of thoughts and experiences, as opposed to being constructed by them. (Albahari, 2010, p.83)

This is only a general approximation of what we naturally take ourselves to be but even in this admittedly preliminary form much has been left out: I do not just take myself to be a thinker or agent; I take myself to be, among many other things, a son, a brother, a friend, a philosophy teacher, and a Bruce Springsteen fan. Albahari's account can in fact accommodate this: we *do* take ourselves to be more than the above definition suggests but such self-descriptions are *grounded* in this more fundamental sense of ourselves as subjects of experience. She identifies the key feature of such a conception as *ownership*. This feature itself has two aspects: perspectival ownership and personal ownership. We have already seen one account of perspectival ownership in Johnston's notion of the "arena of presence" in §5.9 (pp.155-6). It will be illuminating to look at how Albahari also attempts to characterise such a crucial feature. Perspectival ownership, for Albahari, speaks to the sense we have of being the subject of our experiences, as opposed to the objects of which we are aware. This subject is:

> the inner locus of the first-person perspective: the conscious embodied viewpoint from which the world is apprehended. The subject's modus operandi is to observe or *witness* objects through a variety of perceptual and cognitive modalities. (Albahari, 2010, p.83, original italics)

This characterisation is hard to disagree with: although I may not be thinking explicitly about my perspective, in my usual dealings with the world I have the sense that my most intimate and essential being is located somewhere in my head, roughly between my ears and behind my eyes. This sense of position might not be permanently present; there may be moments when I am engaged in some practical endeavor, for example, where a distinction between my body as a whole and myself is not showing up. As Strawson argues, however, this awareness "is fully compatible with our thinking of ourselves primarily or centrally as mental things; and those who stress somatic awareness risk forgetting that it is just as true to say that there is constant background (as well as foreground) awareness of our minds" (Strawson, 1999b, p.129). Even when I do become fully focused on some practical endeavor, the sense of being essentially situated in my head is *readily* slipped back into, and it is especially prominent when I am thinking through some issue, or narrating my experience to myself. One could put the point like this: it is reasonable to suppose that

this implicit sense of my own position would be different if evolution had seen fit to place our eyes on the end of each of our little fingers, and our ears on our chest. Quite what it would be like to occupy the perspective(s) of such a being is hard to imagine, but it seems safe to assume that our experiential sense of position, i.e. where we - in the form of thinking subjects - seem to be located, would be different from where it is now. Had our evolutionary history developed along such lines it seems unlikely that we would still be inclined, for example, to encourage a particularly solemn and quiet friend to become more engaged by saying: "Try to get out of your own head for a bit". It seems likely that such metaphorical advice would not have the intuitive appeal that it has for creatures like us, whose visual and audio inputs are situated on the front and sides of our heads.

Having said as much, our experiential perspective is not limited to only this implied position. Although this is an important part of it, it is not the whole story: when I experience a pain in my foot, for example, I do not experience it as coming from my head. The pain is just *there*: either located in my foot or, if at that moment I feel obliged to consider the phenomenal character of pain as it is in itself, seemingly not located anywhere in particular - there is just a sense of pain occurring *here*, for me. This latter point speaks to the broader sense in which Albahari identifies self-experience as intimately related with perspectival ownership:

> Any conscious creature is uniquely positioned to observe (via witness-consciousness) an array of such objects as pains, thoughts, or its own body, from a perspective to which no other creature has direct access. Insofar as various objects appear to a subject's perspective, in this direct first-personal way, the subject can be termed a *perspectival owner* of other objects. (Albahari, 2010, pp.83-4, original italics)

Perspectival ownership itself, then, has two aspects: I have the sense that I am the owner of a stream of experience; the subject, *here*, as opposed to the "objects" (be they worldly objects or thoughts and sensations), *there*, and I have the sense that the stream I own is this particular one *here*, as opposed to any others out *there*. My claim is that it is this first sense of perspectival ownership that is exposed as inaccurate when we attend to the luminosity of our experiences. Before looking at why this

might be the case we need to consider Albahari's second kind of ownership, as it will become relevant shortly.

Personal ownership speaks to the sense we have of being a subject "with an *identity* (or 'who-I-am-ness') as opposed to a merely impersonal point of view" (Albahari, 2010, p.84, original italics). It is this sense of ownership that is responsible for the difficulty we face in trying to pin down what it is that we naturally take ourselves to be: our identity in *this* sense covers a lot of ground. My social status goes towards my general sense of "who-I-am-ness", as do my autobiographical details, my preferences, and my relationships. My possessions feed into it: my house and my car can be appropriated into a general sense of "who-I-am-ness", to various degrees. It is also not a *static* feature of my self-conception: depending on the social context, varying aspects of my character can come to the fore or recede into the background. My overall sense of "who-I-am" might be quite different if I am socialising with my friends as opposed to being questioned by an authority figure. In my own experience, it is this sense of "identity" that those unfamiliar with the literature take the philosophical problem of personal identity to be dealing with. In discussing personal identity in terms of person *A*'s numerical identity with person *B* at a later time, it has not been uncommon to find first-year students conflating such notions of identity by talking about the *kind* of person they are - their identity in the sense of their preferences and dispositions. To a certain extent, perspectival ownership speaks to how we *experience* ourselves, whereas personal ownership speaks to how we *think*, and talk, about ourselves. This is by no means a clean distinction, and the two feed into each other, but it seems broadly accurate.

Ownership, then, characterises a significant deal of the ego-entity that we naturally take ourselves to be. We now need to look at the role of luminosity in the phenomenological process of recognising our non-identity with such an ego. The crucial upshot of ownership on this issue is that it serves to create the impression that what we are "stands *behind*, and is somewhat in charge of, the stream of thoughts and experiences, as opposed to being constructed by them" (Albahari, 2010, p.83, my italics). As we saw earlier in §5.8 (p.151) this was an objection raised against naïve phenomenalism: we do not typically feel that we are identical with our experiences but, instead, we feel that we "have" them; that we own them. The structure of our conscious experiences is what gives rise to this sense of perspectival ownership, and it does it in at least two ways. Firstly, as we saw detailed by both Johnston (2010) in

§5.9 (pp.155-6) and Albahari (2010, p.83) here, the perspectival presentation of objects results in an implied position. Often this seems to be somewhere in our heads owing to the viewing point implied by the visual presentation of the world as being out *there* - in front of me. This gives rise to the impression that I am located at the centre of this "arena of presence". Although the following explicit reasoning is not involved, it speaks to the effect of such an implied position: if *x* is at the centre of *y*, then *x* cannot be identical with *y* considered as a whole. Ultimately, we have the natural impression that we are in some way distinct from our experiences whilst not being entirely divorced from them either.

The second way that the structure of experience results in such an impression has to do with introspection. In dealing with some particular object in the world, I am aware of it as something that is distinct from me: I can inspect various parts of it and compare it to other objects. We have a fundamental sense of this epistemological relation as a *distinction* between the knower and what is known, which is itself informed by our implied position. The claim is this: when we turn our attention to our experiences themselves, we typically continue to work under this inappropriate model. We can move our attention from one aspect of our experience to another, in just the way that we can attend to one object or another in the external world. This can give rise to the impression that we are looking at our experiences as things distinct from ourselves. In effect, the sense of being an observer, or knower, *distinct* from what is observed, or known, stays with us as we introspect our experiences. As we saw in §2.5 (pp.55-60) this is a mistake: we are not aware of our experiences in the same way that we are aware of objects in the world around us. We can come to see this error through philosophical analysis of the kind we have been conducting. We can also, however, come to see the error experientially, in an immediate fashion and, in so doing, gain an insight into our own nature. This, at least, is the claim under consideration here. Why then should attending to the *luminosity* of experience expose the inaccuracy of this self-conception?

I think this occurs in two ways: attending to luminosity weakens the illusion of being at the *centre* of an experience, whilst also challenging the natural impression that we are aware of experiences in an objectual way. As we saw, luminosity is not a typical phenomenal property: it is not any of the "seen" content but the very "seeing" of it (Fasching, 2010, p.208). It is not, then, the kind of thing we can be made aware of by looking *at* it in the way we look at objects or *seem* to look at experiences when

we introspect. Attending to such an aspect of one's experience might then force the realisation that one is not a *distinct* observer of it.

The relationship between luminosity and our implied position at the centre of our phenomenal field is an equally significant issue. In recognising that luminosity is an intrinsic aspect of the *entire* experience of which I am aware, it can become apparent that what I am (as the being who is aware of this luminosity) cannot *merely* be at the centre. The implied position at the centre of the phenomenal field is, instead, only *part* of the total luminous experience, the entirety of which I am aware. There is a flipping of sorts: instead of regarding my experience as something that I can observe from a central location, this perspective *itself* is seen as only one part of the entire phenomenal scene of which I am aware. It is more accurate to say that it is consciousness that is aware of the "subject", rather than the other way around: this experienced flipping of perspective motivates the term "personal Copernican revolution". Crucially, however, this is not because consciousness is distinct from both the "subject" and object components of the experience (as Fasching holds), but because the self-intimating and self-conscious experience has, as *part* of it, an implied centre. The conscious being recognises itself as the total phenomenal episode, as opposed to an entity that stands apart from it and owns it. In first-personal terms, I recognise the ego-entity that I have been identifying with to be an appearance *within* myself as this self-luminous experience. In metaphorical terms, if I focus on the luminous aspect of my nature as this currently occurring experiential episode, I have the impression of being the light of consciousness: the presence in which all of the content is arising (Fasching, 2010, p.211). Such descriptions of self-experience are common in Indian philosophy: *The Ephemeral Self* account helps to explain why they are: we *are* consciousness, and we can recognise this by attending to a particular aspect of our nature. Śankarā offers a particularly poetic description of his phenomenology when he claims the following: "I am Seeing, pure and by nature changeless … I am unborn, abiding in Myself" (Śankarācārya, 1992, p.123). Being an Advaitin he holds that such "Seeing" is eternal but, taken only as a phenomenological description, it seems to me that such a self-definition could equally well apply to someone who was explicitly attending to the luminous aspect of the experience they are constituted by, without a commitment to the timeless nature of the self. The luminous aspect of experience is "changeless" in that it is constantly present, as a ubiquitous feature of phenomenal consciousness. It is "unborn" in the sense that it

does not move from non-appearance to appearance, such as a particular content does: it is there in each phase of the diachronically co-conscious stream.

§6.6 Death revisited

Having seen through the illusion of being an entity that stands apart from experience, the usual extension of the ego into the area of personal ownership is less likely to occur. This, claim both the Buddhists and Advaitins, will lead to a decrease in unnecessary suffering. Such a claim is too substantive to be investigated in any detail here, but it is tempting to predict that such a loss of emphasis on one's sense of "who-I-am" might well be met with an increase in wellbeing. Globally speaking, the tribal mentality that is in part responsible for many damaging world events is itself largely based on a sense of ownership: this is *my* clan and *my* history. In a less dramatic way, it seems reasonable to suppose that a good deal of personal angst is directly related to the image we have of ourselves, and our natural desire to defend and promote it. However much we might recognise the futility of it, we *do* care what others think of us: attending to our own self-image and presenting it to others in its best possible light is one way that we try to appease this concern. It is obvious, however, that even if such an insight might result in an alleviation of suffering it would do little to help anyone who did not already have their basic survival needs met.

There is, also, the issue of personal death. If we identify ourselves as an ego-entity and extend this sense of ownership to our wider personality and autobiographical history, then death is indeed a terrifying prospect. We may well be able to console ourselves by considering Wittgenstein's (2007, p.106) point that death is not something that *happens* to us, lying as it does outside of our experience. We may be comforted by the notion that what it will be like to be us after death will be exactly the same as what it was like to be us *before* we were born, i.e. nothing to worry about. But death will still be the dissolution of all of the myriad factors that go into what makes my sense of "who-I-am" what it is: my decades-long narrative, my social relationships, my goals and projects, and my memories will all end. If, on the other hand, we identify ourselves as this currently occurring stream of experiences in a genuinely transformative fashion, then the end of the ego might strike us as less terrifying. Perhaps the death of my genuine self, i.e. the end of this stream of

experiences, will also be less concerning. Genuinely, and *experientially*, recognising ourselves as an egoless stream of experiences might well change much of our psychological dispositions, including the fear of death. If I was to genuinely apprehend myself as this stream of experiences by recognising my luminous nature, the desire for tomorrow's stream to be *me* might not arise. This is difficult to know without experiencing such a process personally, but it seems possible. There is admittedly a problem with this defense: until we succeed in carrying out such a process of self-realisation the claims of *The Ephemeral Self* account are somewhat shocking, if the arguments strike us as powerful. An intellectual acceptance of its claims, in the absence of transformative experiential recognition of the self that it describes, leaves us in an unstable middle-ground, psychologically speaking.

There is, however, the following possibility to consider. The luminosity of consciousness is a property of phenomenal content, albeit an unusual one. When your stream of consciousness ends so too does the luminosity of its content. The property of luminosity, however, has not ceased to exist: all streams of consciousness have it. In this sense a crucial aspect of what we are "outlives us". Quite how we spell this out depends on what view of properties we adopt: if properties are universals then the *very same* luminosity of which you partly consist is already in every other stream of experience. It will also exist for as long as conscious beings inhabit the universe. This is close to Advaita's concept of Brahman: the one self of all selves. If, alternatively, we adopt the trope view of properties then it is still the case that all experiences possess a property that exactly resembles an aspect of your nature. Such a thought might be of *some* consolation in thinking about such matters.[36]

The foregoing is, admittedly, a somewhat more speculative series of claims than those that have been put forward in the main body of this thesis. To a certain extent whether or not it is an accurate description of the experiential recognition aimed for by Advaita and Buddhism can only be verified by personal investigation, and I cannot claim any certainty on my part in this respect. Having said as much, it strikes me that such an interpretation is plausible and, crucially, it does not posit self-luminous consciousness as *distinct* from its content, in the way that Fasching does.

[36] It has also been proposed that attending to the luminosity of consciousness can lead to a blissful state in and of itself. Advaitins use the concept of "sat-cit-ānanda", meaning "being-consciousness-bliss", to denote the three aspects of this single state of being. This is an interesting thought but beyond our present scope: I have aimed only to investigate the nature of the first two aspects of such a reality - the possibility of *being consciousness*.

In closing, then, for Fasching, "[the] essence of the meditative process of becoming self-aware is, in my view, a de-identification from what we normally ascribe to ourselves (i.e. what we take to be our "inwardness")" (Fasching, 2008, p.478). This is, of course, a highly radical view: we are neither the biological organisms that produce our conscious experiences, the social self that we typically take ourselves to be or *even*, for Fasching, the experiences that comprise our stream of consciousness. We are simply the witnessing of all that arises for us. *The Ephemeral Self* account, instead, acknowledges that an *aspect* of our nature is such that it can be described as an *atypically phenomenal* "seeing" but denies that this is *all* that we are. It agrees with the general Buddhist refusal to accept the existence of an eternal self, but accepts *genuine* personal identity over time. It can agree with the spirit of Advaita, in holding that a recognition of the luminosity of consciousness can result in a de-identification of sorts, but denies that consciousness is distinct from its content. It therefore represents a middle-ground in this debate. As I have said, the veracity of its soteriological claims can only be tested against personal experience. But the following point seems reasonable: if such radical self-realisation is both possible and successfully undertaken, it *might* in fact change our relationship to death in a beneficial manner.

Conclusion

We began this investigation with a warning from Hume:

> It is certain there is no question in philosophy more abstruse than that concerning identity, and the nature of the uniting principle, which constitutes a person. (Hume, 2007, p.127)

It seems unlikely that the foregoing discussion will do much to dissuade anyone sympathetic to Hume's claim. The claims and positions of the various theories we have covered are each open to counter-arguments from rival views. A gain in theoretical parsimony in area *x* seems often to lead to a loss in intuitive appeal in domain *y*, and vice versa. Measuring the various strengths and weaknesses of any one theory and comparing it with the alternative views of another is a complicated matter, and it is tempting to suppose that a fully convincing account may remain forever out of our grasp. This is, at least, a view that we cannot blame Hume for adopting. This state of affairs applies to many other areas of philosophy as well, of course, owing to the very nature of the questions and methods of which the pursuit is comprised. That being said, I hope to have advanced something of worth in the foregoing discussion.

For one thing, we have seen that the acceptance of just two very plausible and intuitive premises leads to an important, and surprising, account of the relationship we each have to our own experiences. The first premise was that inner awareness is a ubiquitous feature of consciousness. For some this premise has seemed so obviously true as to require no philosophical defence (see §3 (p.80)): it seems enough to point out that an experience that fails to be apprehended by its subject is, in effect, an "unconscious consciousness" and as such is an incoherent postulation. Although, of course, not everyone agrees with such an assessment, the intuitive appeal of the premise, leading to such complacency, is hard to deny. Further to this we saw that the plausibility of a ubiquitous inner awareness (UIA) was strengthened by its ability to cope with The Novelty Problem. To recap briefly: when we attend to the intrinsic phenomenal character of our experience it does not seem to us that something new has been revealed. UIA offers a straightforward explanation for this: we were already

conscious of our experience, and hence nothing new is discovered upon introspection. We saw that other views of inner awareness did not fare as well and, whilst no knockdown argument was offered against the many different ways such accounts could try to overcome The Novelty Problem, UIA's ability to easily, and intuitively, account for the lack of phenomenal novelty gave it a clear advantage. The most serious problem with UIA is arguably the difficulty we face in trying to offer a naturalistic account of it: how exactly does consciousness apprehend itself in such a manner? Taking consciousness seriously, however, and adopting a "consciousness-first" approach (Goff, MS), demands our rejection of the assumption that the correct account of the phenomena will sit well with our reductive ambitions.

The second plausible premise was that experience is self-intimating. The rival accounts of the higher-order theorists were found to be problematic in various ways, some more fundamental than others. The debate concerning higher-order theories is still very much underway in the literature and it would have been foolish to expect a definitive assessment given the scope of this investigation. Having said as much, we identified a number of problems facing such views that do not arise for the self-intimation model. We saw, once again, that the premise under consideration was phenomenologically sound and generally plausible, but not as amenable to a naturalisation of consciousness as its rivals.

Having defended the accuracy of these two plausible, intuitive and widely accepted premises we saw that an important entailment followed from their acceptance - one that seems to have gone unrecognised by many. Succinctly put, if an experience (*x*) is both self-conscious (in the sense that inner awareness is present) and self-intimating, then the only being that can be (non-inferentially) conscious of *x* is *x itself*. If *you* are (non-inferentially) conscious of an experience right now, then it follows that you are exhaustively constituted by that experience for the duration of its occurrence. This is, of course, a tentative claim: it requires the truth of both premises identified above and these matters were not irresistibly settled by any means. It does, however, appear to be both an important and quite simple entailment should those claims hold. It need make no reference, as some similar accounts do, to the notions of "mineness" or "primary presence" in order for the argument to work. It is tempting to guess that the *simplicity* of the argument is partly responsible for its previous lack of articulation in this form. Of course, I cannot claim to have read *all* of the literature that exists on the issue, such work dates back thousands of years across multiple disciplines and

cultures, and it is possible that such an argument has been put forward before (or something very close to it). But it does appear to have been overlooked in the contemporary areas we have covered: for one thing, it seems to me that a naturalistically inclined higher-order theorist could use it in precisely the opposite direction than has been presented here. Noting that an acceptance of self-intimation in conjunction with UIA entails the dubious notion that we are identical with our experiences, he or she might want to employ the argument as a reductio ad absurdum against the self-intimation view. For many of the higher-order theorists it seems quite obvious that we are not, in ourselves, intrinsically conscious beings: we are made conscious in virtue of some of our mental states. To my knowledge no higher-order theorist has used the argument in such a way and this suggests that these entailments have not been fully recognised. We also saw that Dainton's C-system approach was put under pressure by *The Exhaustive Constitution Thesis* and, whilst there are moves that can be made in response, the fact that such a consideration was not dealt with in Dainton's substantially detailed study of the unity of consciousness and its relation to the self suggests that the entailments of The Exclusivity Argument had not been apparent.

We then saw that the findings so far reached, in conjunction with a plausible account of temporal consciousness, led to *The Stream Thesis* and the novel *Ephemeral Self* theory of personal identity. The final move towards this account came from responding to The Bridge Problem by biting the bullet: what we are does *not* survive periods of unconsciousness. In defence of such a move I argued that the various attempts we have seen to overcome The Bridge Problem also result in counterintuitive consequences. Gaps in consciousness pose no problem for virtual or intentional accounts of the self, but the reality of the self is thereby denied: that the survival of myself is a matter of convention is arguably less intuitive than the claims of *The Stream Thesis*. We saw that The Bridge Problem can also be overcome by identifying the self as the substrate of its stream of consciousness, and considered one of the strongest accounts of this kind on offer: Dainton's C-system approach. Once again, the solving of The Bridge Problem came at a cost: the relationship between a subject and its experiences entailed by such a move brought considerable problems. Succinctly put, it seemed to misdescribe the *way* that we, as subjects, are conscious. *The Ephemeral Self* account, in accepting the fatal nature of gaps in consciousness, avoided all of these issues: the reality of the self was fully recognised and the intimate

relationship between a subject and its experiences was kept intact. Noting that any theory of the self will, to a certain extent, clash with our intuitions it was argued that *The Ephemeral Self* is - on balance - a plausible theory. We saw that its counterintuitive features result from our pre-theoretical conception of ourselves, whereas its intuitive appeal was a direct result of its adherence to certain basic phenomenological features of experience. In totalling up the various strengths and weaknesses of a theory such a contrast in the *reasons* for intuitive acceptance or rejection counts very much in *The Ephemeral Self*'s favour: judgements resulting from a careful examination of our immediate experience are likely to be much more reliable than any pre-theoretical notions of the self.

We also saw that the account offered here is arguably of the *moderate* kind when situated in the context of the debate concerning selfhood found in Indian Philosophy. It sat between the Advaitic notion of a life-long (eternal) self and the Buddhist rejection of genuine identity over time. In a more speculative section, we saw that accepting *The Ephemeral Self* might in fact bring soteriological gains: a core claim uniting the disparate philosophies found in the Indian sphere is that suffering can be alleviated through a recognition of our true nature. Such a discovery must, by necessity, result in a revision of our previous understanding of ourselves. An attempt was made to incorporate an interpretation of such self-recognition into the claims of *The Ephemeral Self* and, although there is more work to be done here, a plausible picture was suggested. The theory defended here has, then, a further strength: it offers the possibility of meaningful self-realisation without recourse to anything supernatural or eternal. At the same time it accepts genuine identity over time and as such is more intuitively palatable than the no-self theory of Buddhism.

It seems fitting to end with the following self-appraisal from Hume:

> I had entertain'd some hopes, that however deficient our theory of the intellectual world might be, it wou'd be free from those contradictions, and absurdities, which seem to attend every explication, that human reason can give of the material world. But upon a more strict review of the section concerning personal identity, I find myself involv'd in such a labyrinth, that, I must confess, I neither know how to correct my former opinions, nor how to render them consistent. If this be not a good *general* reason for scepticism, 'tis at least a sufficient one (if I were not already abundantly supply'd) for me to entertain a

> diffidence and modesty in all my decisions. (Hume, 2007, p.399, original italics)

Such a moderate scepticism is wise: each and every account of inner awareness, experience, and personal identity that we have surveyed is vulnerable to objections, and the fundamental assumptions grounding such theories are also very much subject to debate. As such, it would be unwise to expect a definitive account of the relationship between consciousness and selfhood anytime soon. Having said as much, within the landscape of possible options we have here seen another worthy competitor. *The Ephemeral Self* account represents a genuine alternative to its experience-based sister accounts, one that is both grounded in careful phenomenological investigation and conceptual analysis and free from certain counterintuitive consequences (although, as we have seen, not without its own). In the philosophical "labyrinth" that opens up before us in seeking to answer that seemingly simple question "What am I?" we have, then, found one more direction that we can take.

Bibliography

Albahari, M. (2009). Witness-consciousness: its definition, appearance and reality. *Journal of Consciousness Studies, 16*(1), pp.62-84.

Albahari, M. (2010). Nirvana and ownerless consciousness. In M. Siderits, E. Thompson, & D. Zahavi (Eds.), *Self, no self?: perspectives from analytical, phenomenological, and Indian traditions*. Oxford: Oxford University Press. pp.79-113.

Aristotle (1936). *De Anima.* (W. S. Hett, Trans.). Massachusetts: Harvard University Press

Armstrong, D. M. (1981). What is consciousness? In J. Heil (Ed.), *The nature of mind*. New York: Cornell University Press. pp.721-8.

Armstrong, D. M. (1989). *Universals: an opinionated introduction*. Boulder: Westview Press.

Bayne, T. J. & Chalmers, D. J. (2003). What is the unity of consciousness? In A. Cleeremans (Ed.), *The unity of consciousness*. Oxford: Oxford University Press. pp.23-57.

Bayne, T. J. (2004). Self-consciousness and the unity of consciousness. *The Monist, 87*(2), pp.219-236.

Bayne, T. J. & Dainton, B. F. (2005). Consciousness as a guide to personal persistence. *The Australasian Journal of Philosophy, 83*(4), pp.549-71.

Bayne, T. J. (2010). *The unity of consciousness*. Oxford: Oxford University Press.

Block, N. (2011). The higher order approach to consciousness is defunct. *Analysis, 71*(3), pp.419-431.

Broad, C. D. (1923) *Scientific thought*. London: Kegan Paul, Trench and Trubner.

Carruthers, P. (1989). Brute experience. *The Journal of Philosophy, 86*(5), pp.258-269.

Chalmers, D. J. (1996). *The conscious mind: in search of a fundamental theory*. Oxford: Oxford University Press.

Chalmers, D. J. (2010) The singularity: a philosophical analysis. *Journal of Consciousness Studies, 17*(9-10), pp.7-65.

Coleman, S. (2013). The real combination problem: panpsychism, micro-subjects, and emergence. *Erkenntnis, 79*(1), pp.19-44.

Coleman, S. (2015). Quotational higher order thought theory. *Philosophical Studies*. Available at: <link.springer.com/article/10.1007%2Fs11098-015-0441-1> [Accessed 24 August 2015]

Coleman, S. R. (2000). Thought experiments and personal identity. *Philosophical Studies, 98*(1), pp.51-66.

Dainton, B. F. (1996). Survival and experience. *Proceedings of the Aristotelian Society, 96*, pp.17-36.

Dainton, B. F. (2004a). Higher-order consciousness and phenomenal space: reply to Meehan. *Psyche, 10*(1). Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.188.8844&rep=rep1&type=pdf> [Accessed 24 August 2015]

Dainton, B. F. (2004b). The self and the phenomenal. *Ratio, 17*(4), pp.365-389.

Dainton, B. F. (2004c). Unity in the void: Reply to Revonsuo. *Psyche, 10*(1). Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.114.8180&rep=rep1&type=pdf> [Accessed 24 August 2015]

Dainton, B. F. (2006). *Stream of consciousness: unity and continuity in conscious experience* (2nd ed.). New York: Routledge.

Dainton, B. F. (2008a). The experience of time and change. *Philosophy Compass, 3*(4), pp.619-638.

Dainton, B. F. (2008b). *The phenomenal self*. Oxford: Oxford University Press.

Dainton, B. F. (2008c). Sensing change. *Philosophical Issues, 18*(1), pp.362-384.

Dainton, B. F. (2012). Selfhood and the flow of experience. *Grazer Philosophische Studien, 84*(1), pp.161-200.

Deikman, A. J. (1996). "I" = awareness. *Journal of Consciousness Studies, 3*(4), pp.350-6.

Descartes, R. (1986). The second meditation. In: J. Cottingham (Ed.), *Meditations on first philosophy: with selections from the objections and replies*. Cambridge: Cambridge University Press. pp.16-23.

Dennett, D. C. (1978). *Brainstorms*. Sussex: Harvester Press.

Dennett, D. C. (1993). *Consciousness explained*. Harmondsworth: Penguin.

Fasching, W. (2008). Consciousness, self-consciousness, and meditation. *Phenomenology and the Cognitive Sciences, 7*(4), pp.463-483.

Fasching, W. (2009). The mineness of experience. *Continental Philosophy Review, 42*(2), pp.131-148.

Fasching, W. (2010). 'I am of the nature of seeing': phenomenological reflections on the indian notion of witness-consciousness. In M. Siderits, E. Thompson, & D. Zahavi (Eds.), *Self, no self?: perspectives from analytical, phenomenological, and Indian traditions*. Oxford: Oxford University Press. pp.193-216.

Findlay, J. N. (1955). The logic of bewusstseinslagen. *Philosophical Quarterly, 5*(18), pp.57-68.

Giles, J. (1993). The no-self theory: Hume, Buddhism, and personal identity. *Philosophy East and West, 43*(2), pp.175-200.

Goff, P. (2006). Experiences don't sum. *Journal of Consciousness Studies, 13*(10-11), pp.53-61.

Goff, P. (MS). *Consciousness and fundamental reality*. Chapter 1. Available at: <http://www.philipgoffphilosophy.com/uploads/1/4/4/4/14443634/chapter_1_-_putting_consciousness_first1.pdf> [Accessed 24 August 2015]

Hill, J. (2004). The philosophy of sleep. *Richmond Journal of Philosophy (6)*(Spring). Available at: <http://www.richmond-philosophy.net/rjp/back_issues/rjp6_hill.pdf> [Accessed 24 August 2015]

Hoerl, C. (2009). Time and tense in perceptual experience. *Philosophers' Imprint, 9*(12), pp.1-18.

Hume, D. (2007). A treatise of human nature. In D. F. Norton & M. J. Norton (Eds.), *David Hume: a treatise of human nature: a critical edition* (Vol. 1). Oxford: Oxford University Press.

James, W. (1890). *The principles of psychology*. New York: Henry Holt & Co.

Janzen, G. (2006). Phenomenal character as implicit self-awareness. *Journal of Consciousness Studies, 13*(12), pp.44-73.

Johnston, M. (2010). *Surviving death*. Princeton: Princeton University Press.

Johnstone Jr, H. W. (1973). Toward a philosophy of sleep. *Philosophy and Phenomenological Research, 34*(September), pp.73-81.

Kant, I. (2007). *Kritik der reinen vernunft*. (N. K. Smith, Trans.) Critique of Pure Reason: Immanuel Kant (Revised 2nd ed.). New York: Palgrave Macmillan.

Kind, A. (2003). What's so transparent about transparency? *Philosophical Studies, 115*(3), pp.225-244.

Kriegel, U. (2004). Consciousness and self-consciousness. *Monist, 87*(2), pp.182-205.

Kriegel, U. (2009). *Subjective consoicusness: a self-representational theory*. New York: Oxford University Press.

Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly, 64*(October), pp.354-361.

Lewis, D. (1986). *On the plurality of worlds.* Oxford: Blackwell.

Lewis, D. (1993). Many, but almost one. In J. Bacon, K. Campbell & L. Reinhardt (Eds.), *Ontology, casaulity and mind: essays in honour of D. M. Armstrong*. Cambridge: Cambridge University Press. pp.23-37.

Locke, J. (1975). *An essay concerning human understanding*. P. H. Nidditch (Ed.), Oxford: Clarendon Press

Lund, D. H. (2005). *The conscious self: the immaterial center of subjective states*. New York: Humanity Books.

MacKenzie, M. D. (2007). The illumination of consciousness: approaches to self-awareness in the Indian and western traditions. *Philosophy East and West, 57*(1), pp.40-62.

Manson, N. C. (2000). State consciousness and creature consciousness: a real distinction. *Philosophical Psychology, 13*(3), pp.405-410.

McClelland, T. (2014). Affording introspection: an alternative model of inner awareness. *Philosophical Studies, 172*(9). Available at: <link.springer.com/article/10.1007%2Fs11098-014-0421-x> [Accessed 24 August 2015]

McGinn, C. (2004). *Consciousness and its objects*. Oxford: Clarendon.

Metzinger, T. (2003). *Being no one: the self-model theory of subjectivity*. Massachusetts: MIT Press.

Metzinger, T. (2005). Precis: being no-one. *Psyche, 11*(5). Available at: <http://www.theassc.org/files/assc/2608.pdf> [Accessed 24 August 2015]

Metzinger, T. (2009). *The ego tunnel*. New York: Basic Books.

Moore, G. E. (1903). The refutation of idealism. *Mind, 12*(48), pp.433-453.

Nida-Rümelin, M. (2007). Transparency of experience and the perceptual model of phenomenal awareness. *Philosophical Perspectives, 21*(1), pp.429-455.

Olson, E. T. (1998). There is no problem of the self. *Journal of Consciousness Studies, 5*(5-6), pp.645-657.

Olson, E. T. (2007). What are we? *Journal of Consciousness Studies, 14*(5-6), pp.37-55.

Papineau, D. (2002). *Thinking About Consciousness*. Oxford: Oxford University Press.

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

Peijnenburg, J. & Atkinson, D. (2003). When are thought experiments poor ones? *Journal for General Philosophy of Science / Zeitschrift für Allgemeine Wissenschaftstheorie, 34*(2), pp.305-322.

Phillips, I. (2010). Perceiving temporal properities. *European Journal of Philosophy, 18*(2), pp.176-202.

Reid, T. (1854). *Essays on the intellectual powers of man*. (Fifth ed.) J. Walker (Ed.), Derby: Boston.

Revonsuo, A. (2003). The contents of phenomenal consciousness: one relation to rule them all and in the unity bind them. *Psyche, 9*(8). Available at: <www.theassc.org/files/assc/2568.pdf> [Accessed 24 August 2015]

Rosenthal, D. M. (1994). State consciousness and transitive consciousness. *Consciousness and Cognition, 2*(3), pp.355-363.

Rosenthal, D. M. (2012). Higher-order awareness, misrepresentation, and function. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1594), pp.1424-1438.

Russell, B. (1995). *The analysis of mind*. London: Routledge.

Śankarācārya. (1992). Upadeśasāhasrī. In: S. Mayeda (Ed.), *A thousand teachings: the Upadeśasāhasrī of Śankara*. New York: State University of New York Press.

Sartre, J.-P. (2010). *Being and nothingness an essay on phenomenological ontology*. London and New York: Routledge Classics.

Schear, J. K. (2009). Experience and self-consciousness. *Philosophical Studies, 144*(1), pp.95-105.

Schroder, J. (2001). Higher-order thought and naturalist accounts of consciousness. *Journal of Consciousness Studies, 8*(11), pp.27-46.

Searle, J. R. (2000). Consciousness. *Annual Review of Neuroscience, 23*(1), pp.557-8.

Sellars, W. (1963). *Science, perception and reality*. New York: Humanities Press.

Sharma, R. K. (2001). Dreamless sleep and some related philosophical issues. *Philosophy East and West, 51*(2), pp.210-231.

Shoemaker, S. & Swinburne, R. (1984). *Personal identity*. Oxford: Blackwell.

Shoemaker, S. (2003). *Identity, cause, and mind: philosophical essays* (Expanded ed.). Oxford: Clarendon Press.

Siewert, C. (2004). Is experience transparent? *Philosophical Studies, 117*(1-2), pp.15-41.

Simons, D. J. (2000). Attentional capture and inattentional blindness. *Trends in Cognitive Sciences, 4*(4), pp.147-155.

Simons, D. J. & Chabris, C. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception, 28*(9), pp.1059-1074.

Strawson, G. (1999a). The self. In S. Gallagher & J. Schear (Eds.), *Models of the self.* Exeter: Imprint Academic. pp.1-24.

Strawson, G. (1999b). The sense of the self. In C. Crabbe & M. James (Eds.), *From soul to self*. New York: Routledge. pp.126-152.

Strawson, G. (2002). The self and the SESMET. *Journal of Consciousness Studies, 6*(4), pp.99-135.

Strawson, G. (2003). What is the relation between an experience, the subject of the experience, and the content of the experience? *Philosophical Issues, 13*(1), pp.279-315.

Strawson, G. (2006). Realistic monism: why physicalism entails panpsychism. *Journal of Consciousness Studies, 13*(10-11), pp.3-31.

Strawson, G. (2009). *Selves: an essay in revisionary metaphysics*. Oxford: Oxford University Press.

Strawson, G. (2011). Radical self-awareness. In M. Siderits, E. Thompson & D. Zahavi (Eds.), *Self, no self?: perspectives from analytical, phenomenological, and Indian traditions*. Oxford: Oxford University Press. pp.274-307.

Strawson, G. (2013). 'Self-intimation'. *Phenomenology and the Cognitive Sciences, 14*(1). Available at: <http://link.springer.com/article/10.1007%2Fs11097-013-9339-6> [Accessed 24 August 2015]

Strawson, G. (2015). Mind and being: the primacy of panpsychism. Forthcoming in: G. Bruntrup & L. Jaskolla (Eds.), *Panpsychism: philosophical essays*. Oxford: Oxford University Press.

Thomasson, A. L. (2000). After Brentano: a one-level theory of consciousness. *European Journal of Philosophy, 8*(2), pp.190-210.

Thompson, E. (2015). Dreamless sleep, the embodied mind, and consciousness. In T. K. Metzinger & J. M. Windt (Eds.), *Open MIND*. Frankfurt am Main: MIND Group. Available at: <http://open-mind.net/papers/dreamless-sleep-the-embodied-mind-and-consciousness-the-relevance-of-a-classical-indian-debate-to-cognitive-science> [Accessed 24 August 2015]

Tye, M. (1997/2003). Qualia. *Stanford Encyclopedia of Philosophy*, 6; Qualia and Introspection. Available at: <http://plato.stanford.edu/entries/qualia/> [Accessed 24 August 2015]

Tye, M. (2002). Representationalism and the transparency of experience. *Noûs, 36*(1), pp.137-151.

Tye, M. (2003). *Consciousness and persons: unity and identity*. Massachusetts: MIT Press.

Tye, M. (2014). Transparency, qualia realism and representationalism. *Philosophical Studies, 170*(1), pp.39-57.

Warren, H. C. (1957). Buddhism in translations. In S. Radhakrishnan & C. A. Moore (Eds.), *A Source Book in Indian Philosophy* (Vol. 67): Princeton: Princeton University Press. pp.281-284.

Watzl, S. (2011). Attention as structuring of the stream of consciousness. In C. Mole, D. Smithies, & W. Wu (Eds.), *Attention: philosophical and psychological essays*: Oxford: Oxford University Press. pp.145-173.

Wilkes, K. V. (1988). *Real people: personal identity without thought experiments*: Oxford: Oxford University Press.

Williams, B. (1970). The self and the future. *Philosophical Review, 79*(2), pp.161-180.

Wittgenstein, L. (2007). *Tractatus Logico-Philosophicus*. New York: Cosimo, Inc.

Zahavi, D. (1999). Michel Henry and the phenomenology of the invisible. *Continental Philosophy Review, 32*(3), pp.223-240.

Zahavi, D. (2000). Self and consciousness. In D. Zahavi (Ed.), *Exploring the self: philosophical and psychopathological perspectives on self-experience*: Amsterdam: John Benjamins. pp.55-74.

Zahavi, D. (2005a). Being someone. *Psyche, 11*(5), 1-20. Available at: <http://cfs.ku.dk/staff/zahavi-publications/metzinger.pdf> [Accessed 24 August 2015]

Zahavi, D. (2005b). *Subjectivity and selfhood: investigating the first-person perspective*: Massachusetts: MIT Press.

Zahavi, D. (2006). Two takes on a one-level account of consciousness. *Psyche, 12*(2), 1-9. Available at: < http://cfs.ku.dk/staff/zahavi-publications/Two_Takes.pdf> [Accessed 24 August 2015]

Zahavi, D. (2011). Objects and levels: reflections on the relation between time-consciousness and self-consciousness. *Husserl Studies, 27*(1), pp.13-25.