

A Realist Argument for the Self: Emotions and Consciousness in Self-Making

Andrew Oberg

1102354

Supervised by Dr David Morgans

Secondary supervisor Dr Tristan Nash

Submitted in partial fulfillment for the award of the degree of Doctorate of Philosophy

University of Wales, Trinity Saint David

2018/2019

DECLARATION SHEET

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

SignedAndrew Oberg..... (student)

Date06 September 2018.....resubmitted 05 July 2019.....

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used the extent and nature of the correction is clearly marked in a footnote(s). Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

SignedAndrew Oberg..... (student)

Date06 September 2018.....resubmitted 05 July 2019.....

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

SignedAndrew Oberg..... (student)

Date06 September 2018.....resubmitted 05 July 2019.....

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for deposit in the University's digital repository.

SignedAndrew Oberg..... (student)

Date06 September 2018.....resubmitted 05 July 2019.....

Abstract: The question of the self, of what the self is (or even if there is a self) has been one that has grown alongside humanity, one that has haunted humanity, throughout our collective history. It is the purpose of this study to add to that questioning, and to attempt a small contribution to a field that has been as widely covered as it is perplexing. We will undertake this effort by firstly examining some common and representative accounts of the self and what they pertain, and with that as background we will move into the interdisciplinary areas of psychological and neuroscientific concerns regarding the self. We will discover the central role that emotions and intuition play in self formation and function. Applying those lessons philosophically we will build on our (hopefully achieved) foundation and offer a unique definition of the self. Thereby finding phenomenological matters to be of importance, we will next examine two self accounts from those quarters as possible objections to our own, and too conduct a review of phenomenological methodology. Taking that as guide we will explore consciousness and its relation to the self in some depth before finally proposing a metaphysical manner in which the self on our definition might be judged to be realist. With all of the preceding as grounding we will then analyze time for our self-view and suggest that if one's self is to be a personal work – a creation – rather than an accident of happenstance then it is out of the perspective on time whence it will come into fruition. Throughout these necessarily broad but deeply interrelated considerations we will strive to maintain a practical approach and limit ourselves only to the human case.

Keywords: consciousness; emotions; intuition; phenomenology; the self

Acknowledgements: I would like to express my deepest gratitude to my lead supervisor Doctor David Morgans, without whose insightful guidance and consistent challenging this project would not have been possible. I would also like to thank Doctor Tristan Nash for his openness and concern, and my two examiners Doctor Rebekah Humphreys and Doctor Michael Loughlin for their thorough, insightful, and considerate comments. I am indebted to you all. During the course of this undertaking I had the extreme pleasure of presenting at two conferences organized by Doctor Andrea Altobrando at Hokkaido University, Sapporo, and one hosted and run by the Buddhist Studies department at Dongguk University, Seoul. My debt to Doctor Altobrando, Dongguk University, and the many fine participants at each must be articulated as well. At all three events my thinking grew immensely; and the questions from the audience in Seoul were particularly thought-provoking. Finally I would like to humbly recognize the support my wife has given me through these many long years of study. Your patience and steadfastness are astonishing, and in return I can only offer my continued love and mutual backing.

Table of Contents:

Chapter 0: Introductory Remarks: Chapter outlines and key terminology...	1
Section 0.1 A path through the thicket...	1
Section 0.2 Summary of chapter arguments...	5
Section 0.2.A: Chapter 1: Background and literature review...	5
Section 0.2.B: Chapter 2: Psychological and embedded factors, proposal of an alternate self-view...	8
Section 0.2.C: Chapter 3: Phenomenological issues, objections, replies..	9
Section 0.2.D: Chapter 4: Strawson's panpsychism and related issues...	10
Section 0.2.E: Chapter 5: Consciousness and qualia...	11
Section 0.2.F: Chapter 6: The self's 'reality' and self-making...	13
Section 0.3 List of specialized terms and definitions...	14
Chapter 1: Background and Literature Review: Four representative accounts...	19
Section 1.1: Four accounts of the self...	19
Section 1.1.A: The soft anti-realist position...	19
Section 1.1.B: The hard anti-realist position...	25
Section 1.1.C: The soft realist position...	35
Section 1.1.D: The contextualized soft realist position...	47
Section 1.2: Answers? – Guiding questions...	53
Chapter 2: Laying the Groundwork: Psychological and embedded factors, proposing an alternate soft realist self theory...	54
Section 2.1: Situating...	54
Section 2.2: Psychological issues and the self...	55
Section 2.2.A: Cognitive structure...	55
Section 2.2.B: Going back to Kristjánsson's sets...	62
Section 2.2.C: Going back to Hume...	67
Section 2.2.D: Selves, personal identity, and whole persons...	76
Section 2.2.E: Intuitions and emotions...	87
Section 2.3: Bodily and embodied issues...	109
Section 2.4: A word on consciousness and category mistakes...	123
Section 2.5: Certain uncertainty, randomness, and limited choice...	128
Section 2.6: Taking less (but taking)...	140
Chapter 3: Phenomenological Approaches to the Self: Objections, replies, and objections...	144
Section 3.1: Review and preview...	144

Section 3.2: A general objection and reply...	147
Section 3.3: Phenomenologically-based accounts and related issues...	149
Section 3.3.A: Phenomenology 1: Strawson's thin self/minimal subject...	149
Section 3.3.B: Phenomenology 2: Dainton's phenomenal self...	173
Section 3.3.C: Phenomenology 3: Going back to Husserl...	185
Section 3.3.D: Phenomenology 4: Heidegger on the self...	202
Section 3.4 Concluding introduction...	210
Chapter 4: Galen Strawson's Panpsychism, Subjecthood, and the Self...	213
Section 4.1: A starting point...	213
Section 4.2: Panpsychism and subjecthood...	214
Section 4.2.A: Strawson's total physicalism...	214
Section 4.2.B: Responses to Strawson and related panpsychist concerns...	235
Section 4.3: Seductive, but only a shimmering...	251
Chapter 5: Consciousness, Qualia, and the Self...	253
Section 5.1 Modern neuroscience and the structure of consciousness...	253
Section 5.1.A: Consciousness, intent, thinking: Definitions...	253
Section 5.1.B: The gap, the hard problem, the neuronal to experience straits: Navigating...	259
Section 5.1.C: An aside: Always conscious?...	267
Section 5.2: Qualia and the self...	269
Section 5.2.A: 'What it is like'...	269
Section 5.2.B: Artificial intelligence and the self: The importance of qualia...	281
Section 5.2.C: Software, hardware, and wondering about the real...	285
Section 5.3: A query at arrival...	291
Chapter 6: Metaphysics and Time: The Reality of a realist self and its (re-)making...	292
Section 6.1: Metaphysics for a realist self...	292
Section 6.2: Time, the made self, and the making self...	307
Section 6.3: Exhausted conclusion...	328
Bibliography...(334)	

Chapter 0: Introductory Remarks: Chapter outlines and key terminology

0.1 A path through the thicket

What exactly the self 'is' could well be an unsolvable problem. Human societies have after all been pondering and wondering about the self more or less since we started pondering and wondering. Yet if so then why bother devoting our time to it? Why is this particular quest one that we seem incapable of abandoning? A puzzle that we cannot put away, a mystery that we cannot lay to rest? Surely there are more pressing problems that warrant our attention, problems that actually have some hope of being settled. Yet the question comes back, and will not leave us in peace, for what could be more pressing than the quandary of who we are? Of what we are? Just what is this 'me'?

Despite its probable intractability I think that the question of the self is one well worth pursuing due, if nothing else, to its very long reach into the rest of our lives. For reasons that I hope will become clear in the course of our undertaking, I believe that the concept of self which we choose to hold, which we purposely adopt (maybe even take on faith), has the potential to extend itself quite deeply into the meaning and purpose of our lives. I will propose that the self-view one takes has the potential to fundamentally alter how one perceives the world, one's place in it, and one's movement and understanding within that. By offering an alternative self concept that, I believe, is more theoretically sound and more in line with what recent advances in psychology and the cognitive sciences have taught us about how our brains work and our behavior is generated, I wish to lay the basis for where such an account might take us in application and in practice – in short: in living life.

Our study will accordingly be divided into two main parts: the first, and by far the bulk,

will attempt to establish the self concept that I wish to argue for – this is the *what* of the self – and the second will then seek to establish what such conclusions might mean on a more experiential, day-to-day basis – this is the *how* of the self. In the following first full chapter (Chapter 1: ‘Background and Literature Review’) we will cover some of the relevant literature and explore four representative self accounts in order to provide the necessary foundation, and then in Chapter 2 (‘Laying the Groundwork’) we will turn to an examination of contemporary empirical findings so that we may draw on their data before offering our own self theory with a consideration of relevant mind and body issues. In Chapter 3 (‘Phenomenological Approaches to the Self’) we will begin by briefly looking at a potential objection to the self theory introduced in Chapter 2 and thereafter analyze a number of phenomenological perspectives and problems regarding the self. Chapters 4 and 5 (‘Galen Strawson’s Panpsychism, Subjecthood, and the Self’ and ‘Consciousness, Qualia, and the Self’, respectively) will delve into the controversies surrounding the topics of consciousness and mind and their intertwinement with the self, and lastly in Chapter 6 (‘Metaphysics and Time’) we will suggest a metaphysical grounding for the self theory that our study advances and propose how one’s approach to time might prove an important key if we are to take control of the way we are, if we are to shape our existence with an awareness and a purpose.

These concerns seem inseparable, and thus for millennia the received wisdom that our true selves, the real ‘me’ which relates to the wider world, was considered an immortal soul of some kind, an essence that transcended the physical and either went somewhere after death or returned, usually following a time spent in sojourn, in another form to live physically once more; outlooks on which of the two views was preferred were typically culturally determined. René Descartes famously changed the emphasis of this ancient idea by shifting the focus from the ‘soul’ to the ‘mind’, but the general picture of a mysterious driver in a

temporal bodily machine did not change all that much. Descartes' conception itself was considerably more complex than this, however:¹ to him one's true self was the thinking thing, a thing which must be substantive as the acts of thought have to be based *in* something, and that substance was an intellectual substance, the pure intellect.² As Gary Hatfield puts it, Descartes 'equated *soul* with *mind* and preferred the latter term to avoid ambiguity.'³ In the Cartesian way of thinking mind and body are wholly separate entities but clearly interact as is evidenced by sensations like pain or hunger being felt by 'me' and not merely observed by 'me'. The nature of this interaction, the manner in which the two communicated, and how the mind was able to move the body, is not clear in Descartes' account, and he was unable to adequately answer objections of this kind that were raised by contemporaries.⁴

While such are thought-provoking, and while Descartes' replies too are of interest, it is not our purpose to dwell on the Cartesian self; rather we will instead take his ideas as a common currency circulating amongst modern academic efforts and thus we have only very briefly presented them here. Nevertheless, what we have is enough to get started. Over the years since Descartes a number of other philosophical ideas on the self have also been developed that discount or bracket the notion that the self is a 'something else', whatever that may be, dwelling within or alongside our physical bodies. To therefore speak philosophically of a self in the present climate requires an extended further discussion enunciating just what is meant by the term, what version one wishes to endorse or to argue against. We too will follow in the tradition of providing this.

¹ Descartes was the first major thinker to use the term 'mind' instead of 'soul' and saw the relation between the mind and body not as a driver/pilot and vessel but one where 'they systematically affect each other, but not other things, in ways that make the two of them together function as if they were one.' John Barresi and Raymond Martin, 'History as Prologue: Western Theories of the Self' in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), pp. 33-56 (p. 39).

² Gary Hatfield, *The Routledge Guidebook to Descartes' Meditations* (Abingdon, OX: Routledge, 2014).

³ *ibid.*, p. 124; emphases in the original.

⁴ *ibid.*

In the below our study will begin with an exploration of four common theoretical accounts of the self, four main threads into which the primary self concepts in others' works may be tied, examining each for philosophical robustness and correspondence to recent research in cognitive and psychological studies. In that, an aside at this point since a word on the latter seems appropriate: although our central concerns are philosophical, whether or not – and if so to what extent – a self concept aligns with the empirical results of cognitive and psychological research is important because if we hope to arrive at an account of the self that can be applied practically (that is, that can be used more effectively towards meditated self-goals and experiential concerns) then a more solid grounding in how we are built to work (as it were) would be of clear benefit. At the very least it seems reasonable to see what the sciences have to offer regarding our natural functioning when forming a theory of what our selves may be.

Thus the proceeding background literature chapter will be structured by first presenting each of the four views followed immediately by a critique of the self concept concerned, done separately by subsection. The four accounts to be considered have been chosen for their representational qualities and are not meant to be an exhaustive description of the work on the self heretofore. Largely following Kristján Kristjánsson's categories in *The Self and Its Emotions*,⁵ the first account will consider the soft anti-realist position, the second will cover the hard anti-realist position, and the third the soft realist position. The final account will be of the contextualized (or conditioned) soft realist position, a category that Kristjánsson did not consider but which I think is illustrative of a different understanding of the soft realist position and which also brings in situational elements that will be important for our later chapters. The hard realist position, that of a separately

⁵ Kristján Kristjánsson, *The Self and Its Emotions* (Cambridge: Cambridge University Press, 2010).

existing and (usually) eternal Cartesian ego or soul comprising the self and temporarily housed in the body, will not be examined due to its tendency to preempt further debate on the topic and its already common extensions; we know how the hard realist self has historically been implemented, and since we are attempting to achieve something new – or at least varied – by our study of an alternate self concept we will table the familiar.

Following this summary a proposal for another version of the soft-realist self will be offered in Chapter 2 that is loosely based on Kristjánsson's idea of an emotionally grounded self. It will be argued that the proffered view would be more philosophically robust than Kristjánsson's account, better able to deal with likely criticisms of a soft realist conception and issues that arise in conjunction with the self (e.g. body/embodyed issues), and closer in line with the interdisciplinary research on the decision-making and behavior generating traits that are presented in the survey of existing views. This self concept may be labeled the (admittedly somewhat awkward) 'limited choice soft realist self'. With this self theory we will then continue our study as outlined above (and see the section here below for more details on the arguments of each chapter), and throughout we will strive to maintain a focus on the pragmatic and the lived. To study the self, after all, is to study being in action. What consequences might the adoption of such a view as ours have for bodily understanding? How would it situate the self socially? What contributions could an account like the one argued for grant to self-knowledge and self-creation? These and other inquiries will hold our attention.

Section 0.2 Summary of chapter arguments

A. Chapter 1: Background and literature review

As outlined above, this chapter is intended to present representative accounts of the primary self theories that are currently circulating in the academic literature. The chapter is

not an extensive listing of these theories, nor is it meant to be such. Rather it focuses on a series of single self-views that I believe amply demonstrate the main tenants of each of the four broad categories into which the theories have been placed.⁶ The first, the ‘soft anti-realist position’ is characterized by its focus on individuality and uniqueness without making any subsequent substantive claims for selfhood like those found in Cartesian theories. This account does not differentiate between a self and a person and concerns itself with ‘ownership’ relations, e.g. between a mind and a body, or an individual and her behavior. The self is the ultimate ‘owner’ in such relational strings, but is never clearly defined in a way that is not arguably circular or excessively vague. This aspect of the account is critiqued, as is its failure to work clearly with the nuances involved in mind versus brain issues. The absence of a boundary between a ‘self’ and a ‘person’ is also faulted, and some related concerns of embodiedness, narrative accounts, and embeddedness are considered.

The second self position is that of the ‘hard anti-realist position’, a self-view designed to correct the deficiencies of the ‘soft anti-realist position’ by definitively eliminating any notion of a core self ‘thing’ or ‘substance’. Sometimes also called the ‘reductionist view’ this self theory argues that only because we typically enjoy both physical and psychological continuity do we think there is an ongoing self, but that in fact such is an illusion. Moreover, because there is no real self, ‘you’ may have any number of ‘lives’ within a standard period of biological existence granting that there is a form (any form) of psychological connectedness and/or continuity. The stance taken towards personal identity in this account is criticized, with some added thoughts towards uniqueness covered somewhat heavily. Added to these are concerns of intuitive opposition, moral responsibility, and the failure to consider emergent properties.

⁶ Again, the categorization scheme followed mainly comes from Kristjánsson’s work, see *ibid.*

The third position we examine is the ‘soft realist position’. This self theory asserts a Humean origin and on it the self is described as a ‘psychological unit of reference’ and given a reality that is based on personal emotion. Thus the self is that which meets the image invoked when one speaks of oneself, an image held by oneself and by others. It differs from something like the ‘hard’ realism of a Cartesian ego in that no claims to a ‘resident’ within the body are made, rather that the body itself and its unique sense of identity (including emotional function) are the constitutive self elements. We find this view to have the most merit of those outlined in this opening chapter, but it is nevertheless deemed insufficient for a number of reasons. Its Humean basis can be doubted, but more than that its grounding for the application of ‘realism’ is not justifiably executed. The role of agency and decision-making here also seem to be based on an understanding of human psychology that does not match the results of research in that field, and the functioning of intuition as opposed to emotion is also not adequately handled in our judgment.

The ‘contextualized soft realist position’ is the final self concept presented in our review of the relevant literature. This account too, like that immediately above, posits a substantively real self based in the emotions but seeks to ground it in a neuroplastic model in the operations of a mammalian brain (human and nonhuman mammals). It is the ‘point of view of survival’ and a biological adaptation. This view heavily emphasizes the embedded conditions under which organisms dwell and places agency at that level of surroundings – individuals are so engrossed in their environments and biochemical inheritances and influences that each one cannot help but to do whatever it is they end up doing. Nevertheless moral responsibility for demonstrated acts is still assigned as being co-creative of the ‘me’ associated with said behavior. While the stressing of embeddedness on this model is positively noteworthy, it is critiqued as being possibly overdone, and

furthermore the foundation in neuroplasticity might be stretched as well given that the theoretical and empirical results in the cognitive sciences regarding its conclusions are still very much controversial. The root mental model here could be problematic too.

B. Chapter 2: Psychological and embedded factors, proposal of an alternate self-view

This chapter begins with an overview of contemporary findings in psychological research that point to a conception of mind that is two-tiered, with the root level being characterized by the automatic processing of elements such as intuitive judgments, choices, and biases, as well as emotional reactions. These factors then feed into rational processes, which may or may not play a further role in subsequent decisions. In this way, according to the model, all input both external and internal to (and from) the non-brain portions of the body are processed primarily by the brain in the absence of thinking awareness, and the results of those procedures are then sometimes also analyzed via active thought by organisms capable (to whatever degree) of such. The following focuses entirely on the human case, but naturally many nonhuman animals are able to do so as well, depending on the qualities of the species broadly and the individual within that grouping. The model presented, although tiered, is not a higher-order theory of mind, nor does the argument presented in this study endorse a higher-order conceptualization. Moreover, while intuitions are generated automatically and their output always inclusive in any (potential) subsequent analysis achieved via active and aware thinking, intuitive reactions can themselves be adjusted, changed, or even transformed, and this chapter also covers how such is considered to occur or, if desired, so sought by an individual. (This methodology is returned to in the final chapter wherein self-making is discussed.)

Hume's arguments on the self are re-considered and an alternative interpretation to that

upon which the soft-realist position introduced in the first chapter is based is also conducted, and thereafter the manner in which our own argument segregates into interconnected levels the self, personal identity, and whole persons is given (we label ours the ‘limited choice soft realist self’; on these terms see also the definitions listed in Section 0.3 below). As a part of this intuitions and emotions are more fully analyzed and differentiated (see again the terminology list below, but also Section 5.2 ‘Qualia and the self’), and bodily/embodyed issues. The structure of human consciousness and connected category mistakes in discussions of it is briefly returned to as part of the issues relevant here (see the terms below), but a fuller treatment of consciousness is saved for Chapter 5, particularly the first section (‘Modern neuroscience and the structure of consciousness’). Contextualized factors belonging to one’s fact of embeddedness are also outlined – including randomness and probability – and the resulting reduction of real choice in life that these engender is applied towards the working definition of the self that was given earlier in the chapter.

C. Chapter 3: Phenomenological issues, objections, replies

Chapter 3 begins with a general review of the material covered thus far and importantly recaps the definitions of the self, personal identity, and whole persons upon which the primary movement of the study is based. In response a potential objection to the self as has been presented is considered and a reply given in an attempt to grant additional theoretical strength to the position. The remainder is then spent on phenomenological issues; and this because such concerns are foundational both to our own position taken on the self and to those other positions which might constitute arguments against or alternatives to our view.

Of these Galen Strawson’s thin self or minimal subject is first analyzed in full and rebuttals made to it, focusing on how it fails to satisfactorily account for diachronicity and takes an

uneven (perhaps contradictory) approach to evidence from related areas in the empirical sciences. Strawson's self too, as fully mental, cuts off the bodily concerns which we earlier argued (in Chapter 2) to be crucial, and finally we object to his rejection of emergence which, although it may appear to fit with his own self-view on the face of things, does not match the broader implications involved in what he is proposing without lapsing into questions of regress. Following Strawson's thin self we move to Barry Dainton's phenomenal self, introducing its tenets as we did for Strawson's and then examining it for weaknesses and points at which a critique may be made. Here again we are troubled by an outright dismissal of the body and by what we argue to be a lack of justificatory substance to account for human experience, even purely on psychological levels. We also find confusion related to mind issues versus brain issues, and too for self versus subject.

The chapter closes with a return to Edmund Husserl's exposition of the phenomenological school he created and the further work that his student Martin Heidegger did in the area, applying both thinkers' ideas to the self generally and to our own version of it in an attempt to provide valuable theoretical grounding for the deeper aspects of our study that will follow in subsequent chapters. Of particular note in these sections are Husserl's 'horizon' and Heidegger's 'world', both of which play into the contextualized and embedded concerns we have already highlighted (especially in Sections 1.1.D 'The contextualized soft realist position' and 2.5 'Certain uncertainty, randomness, and limited choice'). Some of Heidegger's ideas on ontological issues are also presented, and these will prove informative for our metaphysical concerns – particularly regarding the 'reality' of our soft realist self – in Chapter 6.

D. Chapter 4: Strawson's panpsychism and related issues

As consciousness was found to be an important definitional element in the alternative

self-view we argued for in Chapter 2, and as concerns related to consciousness repeatedly came to the fore in our considerations in Chapter 3, the topic is more fully focused on in this chapter via an in-depth analysis of Galen Strawson's arguments for panpsychism, chosen because it directly links with his self concept presented in the previous chapter. The objective is therefore to examine the contemporaneously significant topic of the question of panpsychism and the subsequent implications for the self and subjecthood. Strawson's version of panpsychism is rooted in his 'total physicalism', and we open the discussion with an extended explanation of his conceptual framework and the reasons he gives for it. Along the way some questions are raised and some replies that Strawson may make are considered, based on his account and the justifications he outlines. A number of thinkers have responded directly to Strawson's work on this subject, and following the exposition of Strawson's assertions these too are analyzed in turn. Throughout what is sought is a clearer view of what consciousness might be and what place it might hold, not only within human and nonhuman creatures but without them as well. Could consciousness actually form a fundamental part of the universe at large? We conclude that although the notion is an attractive one we do not, at least for the time being, believe that there is sufficient evidence and/or theoretical reasons for thinking so. Nevertheless the question is left open, and we also find that the tendency towards the development of consciousness may be what is really fundamental, and if so then emergence could provide a good explanation for how.

E. Chapter 5: Consciousness and qualia

Chapter 5 continues the exploration of consciousness that was begun in Chapter 2 and has continued to the present point in the study. Having in the preceding chapter dwelt on abstract and metaphysical issues connected with panpsychism, we open this chapter with a return to the cognitive sciences and the empirical grounding that researchers have been able to achieve with regard to human consciousness. Initially some definitions are worked

out for consciousness, intent, and thinking in order to importantly distinguish between terms that are easily confused, a problem that is not helped by the tendency amongst some writers to use such synonymously. Above all, clarity is sought here.

Having (we hope) achieved that, a turn is made to what has been called the ‘hard problem’ of the perceived disconnect between the biological level of neurons and their functioning and the experiential level of the felt and the thought. This is the issue of ‘qualia’, or the ‘what it is like’ aspect of experience, and it is deeply intertwined with much of the work also being done in mind/brain areas. Essentially we take the stance that most of the difficulties arising are a result of conceptual confusion fueled by the inadequacies of language. The root problem, we assert, is that although we are naturally programmed to think and speak (and therefore think again – thought is most often tied in with its linguistic expressiveness) on the abstract level of ideas, our qualia are products of how neuronal mental representations function, processes that are entirely physical and biochemical, making the shift from notional thinking to purely biological thinking cumbersome and counterintuitive. Some reasons related to the brain’s emotional and intuitive operations are expounded (again, on these see the definitions list below), and this once more leads us into a broader discussion of the place of consciousness. We argue that there is good theoretical grounding, and moreover provocative empirical bases, for concluding that consciousness – as a network – ought to be considered always present, even during periods such as blackouts or deep sleep.

The chapter then progresses to a more in-depth consideration of qualia as applied to the self, starting with a review of Thomas Nagel’s initial raising of the issue of qualia and what he has commonly been taken to have meant by it, and thereafter some objections to this interpretation followed by our own alternative view are given. With this new appreciation

for qualia in place we examine the increasingly urgent topic of Artificial Intelligence and whether or not such machines could ever be considered to have their own selves. From there another modern recurring motif is analyzed: the analogy of software and hardware with the self and body/person (the latter depending on if a realist/anti-realist position is taken). This opens our work on metaphysics, reality, and what the 'real' might be.

F. Chapter 6: The self's 'reality' and self-making

Our final chapter focuses more fully on the notion of 'real' that we wish to promote in regards to the self, and specifically with relation to our own 'limited choice soft realist' self-view. An argument is made for a form of reality that is based in function, whereby the effects or results of an abstraction are taken as sufficient anchoring for the kind of realism that we more commonly associate with physical presence. This is not to give the self a physicality, nor a 'hardness' in something like a Cartesian sense, but it is nevertheless to grant substance to the self through the influence that it wields in the life of its bearer/creator/shaper. The self, on our view, is a constant work in progress, both in how it affects the whole person and in how the whole person affects it (on this circularity see Chapter 2, especially Section 2.2.D 'Selves, personal identity, and whole persons').

Due to this ability to actively transform and mould one's self Chapter 6 also digresses into a brief excursion into time, to what it might mean for us in a phenomenological sense regardless of what it may or may not 'be' in a cosmological sense (this latter being a deep point of contention not only in philosophy but in theoretical physics as well). We review some concerns on time from selected phenomenological thinkers on the subject and then take their thoughts into our own on the making of a self as we have come to understand it. This involves one's view on one's own being in time, but rather than presenting a linear or 'moving through' position in this aspect we argue for a 'spiraling' view that takes the self

in time sensationally and placed in an ever-now. Finally a summary of the study is provided, and the primary conclusions arrived at are reviewed and highlighted. There can probably never be an ‘answer’ to the question of the self, but we hope to have provided some additional possibilities along the way.

Section 0.3 List of specialized terms and definitions

A number of specialized terms are included in the following study, and although great pains have been taken to illustrate the usage in each case and how, when applicable, the employed term differs from how it might more commonly be used in either a vernacular or academic sense, in order to try and preemptively prevent confusion on the reader’s part we will list and define these terms here. Included with each are cross-references for sections within the text’s main body wherein further discussion occurs.

Preconscious/conscious: The study uses ‘preconscious’ to indicate the automatic workings of the brain that are connected to data and other input processing and that do not enter directly into noticed awareness. These results do, however, influence aware thought in the pre-reflective (or non-reflective) judgments, biases, influences, effects, et cetera that they carry into the brain’s abilities for subsequent rational analysis. On the two-tiered model of brain function presented here this is the core level. In some ways the label ‘preconscious’ is less than ideal as the study also argues for consciousness as being ongoing and always present (see below), but since the vast majority of writers both past and present have used the term ‘conscious’ synonymously with ‘awareness’ we chose ‘preconscious’ in an attempt to subvert such presumptive nuances. For similar reasons of historical baggage and undesired assumptions this study also avoids ‘unconscious’ and ‘subconscious’. What is meant then by ‘preconscious’ are those mental workings whose output is accessible but whose machinery is (and will remain barring evolutionary changes) always hidden. See

Section 2.2.A ‘Cognitive structure’ and Footnote 46 in Chapter 1 for more details.

‘Conscious’ is used in the following in reference to the functioning of the consciousness network itself (see the next pairing of specialized terms listed below). The term is never used to indicate ‘aware’ except in the case where others have used it that way and their works are being quoted; in all such instances this is pointed out, and this study’s use of ‘aware’ is also indicated when applicable as an instance where others might use ‘conscious’ instead. As explained above, ‘preconscious’ does not collapse into ‘conscious’ because the preconscious cannot be made actively aware, although it is a part of the same overall system and its effects are certainly felt within that system. The prefix ‘pre’ here is less than perfect since it is not technically ‘before’ consciousness, but we have had to compromise with it to avoid an overly cumbersome phrasing (e.g. ‘impossible to be rationally aware of yet still influential on the rational’) or the laden terms ‘unconscious’ and ‘subconscious’. See again Section 2.2.A ‘Cognitive structure’ and Section 2.4 ‘A word on consciousness and category mistakes’.

Awareness/consciousness: ‘Awareness’ is used to indicate accessibility to rational thought, the possibility that a datum could be actively analyzed through mental processes. Again, this differs from ‘preconscious’ analysis in that it is not automatically done by the brain but rather is purposive and determined. If an individual must decide to examine an input or perception (be it of any nature or kind), then that is carried out in a state of ‘awareness’. It should be clear how this aligns with the way many have used ‘consciousness’ – almost in some cases in the sense of merely being awake or paying attention – but we wish to differentiate that, as will now be explained.

On our account ‘consciousness’ refers to the brain’s networked functioning, which at

present cognitive scientists have for the most part (though not entirely, some debates remain) determined to be modular in structure. This amounts to a ‘constellation’ of consciousnesses, with the brain’s communicative pathways providing what is perceived as a sense of unity (a unified consciousness) by us when everything is working the way it should be. The brain may or may not contain an ‘interpreter’ module, which, if it did exist, would exert the role of collating and streamlining the various products of the many other modules. Due to this varied framework it is argued that consciousness should be considered to always be present – even in the absence of awareness – if at least one of the brain’s modules in the network is functioning, and there are good reasons to think that for the entirety of life this is the case. Thus, consciousness is only lost at death if we are correct in our interpretation. Consciousness is discussed throughout the study, but see especially the following: Section 1.2.A ‘The soft anti-realist position’, Section 2.2.A: ‘Cognitive structure’, Section 2.4: ‘A word on consciousness and category mistakes’, Section 5.1 ‘Modern neuroscience and the structure of consciousness’, and Section 5.1.C: ‘An aside: Always conscious?’.

Self/personal identity/whole person: What the self may be thought to ‘be’, the ‘what’ of the question of the self, naturally forms the core of the study. As such this is thoroughly defined in the below a number of times, but here at the start we can indicate that we will place these three terms in a stacked but circular relationship. The self will be argued to be the core concept, consideration, internal view, mental ‘unit’, that a person has in regards to their being. The elements of this are quite bare. Adding to it other elements, also kept to a minimum, brings out the level of personal identity; notable here is the denser application of physical and other contingent details. Finally, rounding out a human subject alive and moving within an environment, is the level of the whole person. Moreover at this highest level, the one at which we go through every moment of our daily routines, lies the

capability to purposefully exert influences on one's own core conceptual self and thereby to adjust it, shifting in turn the resulting levels of personal identity and whole person and the varying related aspects. To make this clearer 'equations' are given in the study, and we will reproduce them here:

- 1) Self = self-constituting emotions + self-comparative and self-conscious emotions + consciousness + bodily presence (merely having a body without further details on that)
- 2) Personal identity = the self (as above) + contingent facts about the body + other contingent facts (i.e. those not about the body proper) + feedback from the social realm
- 3) Whole person = the self (as above) + personal identity (as above) + embeddedness (i.e. all of the environmental, situational, conditional, et cetera details of an enworlded being)

These terms are defined and collected together most fully in Section 2.2.D: 'Selves, personal identity, and whole persons', but of course discussion on the self and related personhood concerns takes place throughout our study.

Emotions/intuitions: The terms 'emotion' and 'intuition' are used in their academic ways in the below, but the position that such hold in the mental model presented may be new or noteworthy to readers without a background in some of the recent advances made in neurological studies. As might be expected, emotional reactions stem from very primitive areas within the brain and have proved to be very useful from a biological point of view in the stimulation of advantageous behavior. This includes, but goes far beyond, reactions such as the 'fight or flight' mechanism. The brain's automatic processing, mentioned above in relation to our use of 'preconscious', generates emotions as a means of organism guidance and processing input.

Similarly, intuitions are also produced by the brain via its preconscious analyses and are used to influence the form of resultant actions taken by the individual. Intuitions differ from emotions in that such are judgments, determinations, biases, prejudices, conclusions rather than feelings. The two systems are similar and it is probable that they often – or even nearly always – work together, but evolutionarily emotions are likely the older of the two, and perhaps the (nuanced) difference might be found to be akin to that between a ‘nudging’ and a ‘prompting’, emotions falling into the previous category and intuitions into the latter. Intuitions, furthermore, are based in knowledge and can be adjusted either through feedback from others in the milieu in which one moves or through a goal directed undertaking that one involves oneself in with the express desire to automatically judge differently or transform one’s assumptions and/or presumptions. It is possible for the two to overlap expressively (we may ‘feel’ them as the same at times), but procedurally and functionally emotions and intuitions are not the same. A great deal of psychological research has been done in the area of intuitions and intuitive decision-making in recent years, and this is covered in the text. See to some degree Section 1.1.C ‘The soft realist position’, Section 1.1.D ‘The contextualized soft realist position’, Section 5.2 ‘Qualia and the self’, but most especially Section 2.2.E: ‘Intuitions and emotions’.

Chapter 1: Background and Literature Review: Four representative accounts

1.1 Four accounts of the self

A. The soft anti-realist position

The first of the four representative accounts of the self that we will examine is the soft anti-realist view, and for it and the others in each case we will proceed by presenting only a general outline of core ideas regarding what the self is held to be (or is not) on the account, and then thereafter offering a short critique. Although the reader may find problems with some or all of the self-views as they are being initially enunciated, a discussion of such will not proceed until the text following the introductions; the first portion is simply meant to touch on the main concepts. Our renderings will be brief and broad pictures.

The soft anti-realist view is the concept of the self that many of us hold without really thinking about it. This position is notable in that it does not distinguish clearly between a ‘self’ and a ‘person’; here the self is seen as a unique individual but without the further Cartesian ego/soul claims being attached. We can find this self concept particularly visible in the background of legal structures such as bills of rights that bestow certain privileges and offer particular protections from the perspective of an individual person living within a society rather than of an individual who is a component part of a society.⁷ Focused fundamentally on the individual and her relationship with the greater community only in so far as she needs shielding from it and liberties to operate as she pleases in defined ways

⁷ That is, the individual’s concerns are given primary importance; e.g. freedom of speech, which may well lead to negative social outcomes in certain instances of its use but which is thought to be of sufficient importance for the individual that it is guaranteed. Framers of such laws naturally see a society in which each individual has free speech as being a better society, but their focus in crafting such a decree is on how an *individual* will live, not on how the society as a whole (in a transcendental sense) will live.

within it, this account takes form through the liberal policies and governance it enacts. Thus it is that Ronald Dworkin takes the essence of liberalism to be equal consideration of and respect for each individual, whose civil liberties are needed to guard them against the preferences of others concerning how they should act.⁸ Similarly, Jan Narveson takes his ‘liberty rule’ to be of paramount importance, by which one may do whatever one pleases granting that such does not harm another or interfere with another’s affairs. This attention to a singular person’s actions and how the actions of others may affect said person rests upon the extension of ownership between that owning and that owned; this is the self, understood here as one’s will or one’s mind, owning one’s body and behavior in the way a material object is owned.⁹ Responsibility for what is done and the resultant consequences therefore lies with each essentially independent mind (= will, = self, = person), and the ramifications of any individual’s acts extend only out to a single degree of separation from their source. According to this view, the self is atomistic and is the ‘owner’ of the activities that are meant to be protected from incursion upon or held responsible for the effects of, depending on the situation. It is thought to accordingly follow that since bodies are owned by minds and are the means by which actions are executed, the chain of ownership-responsibility is: act → physical body that committed act → mind that directed physical body. The self, to put it somewhat crudely, is that (legally) determinable point at which the buck stops.

This account can at first seem reasonable given the background legal environments and educational foci of contemporary liberal societies with which many readers will be familiar (e.g. an emphasis on being one’s own person, standing on one’s own two feet, looking out for number one, et cetera.), but when examined in greater depth a number of problems

⁸ Gerald F. Gaus, *Political Concepts and Political Theories* (Oxford: Westview Press, 2000).

⁹ Jan Narveson, ‘Libertarianism vs. Marxism: Reflections on G. A. Cohen’s “Self-Ownership, Freedom and Equality”’, *The Journal of Ethics*, 2:1 (1998), 1-26.

come to the fore. Primary amongst these is that by this line of reasoning we still do not know what the self is. Is it an ‘individual’ understood as a collection of organic materials animated by a directing mind? Is it that directing mind itself? What exactly *is* a mind? Philosophers and neurologists have long discussed what a mind, as opposed to a brain, can be said to be and have still not arrived at a satisfactory definition.¹⁰ Narveson, moreover, only adds to our confusion in his argument on body ownership, claiming that ‘Everyone is “boss” over his own mind’.¹¹ Yet what might be meant by the ‘one’ in this ‘everyone’? If the self is the mind, then what could be over the mind? Am ‘I’ the boss over my mind? For in that case I (the self) would have to be both my mind and the boss of my mind. What though could possibly stand outside the mind so that it might be its boss if the self is purely and fully the mind? What room is left in this relationship for any form of an external ‘I’? With nothing definitionally left to work with, what might its contents be? All we have really done here is to come back to our original query of what the self is. That there is no substantive self that can be pinned down with any accuracy, but there is still nevertheless a ‘something’ present (and a ‘something’, moreover, that ostensibly holds responsibility and that can be held accountable – legally and morally – for actions and consequences) is why this account is called the ‘soft’ anti-realist position rather than its hard variant.

If, on the other hand, the self by this view is not understood as the mind proper but instead as being like the full individual (in the dominantly legal sense), then we encounter a further difficulty, namely, that since there is no distinction drawn between the ‘self’ and the ‘person’ we must assume that an ‘individual’ more or less equates to a ‘person’. Could this be extended further to equate with the self? Not, I think, without either embracing a non-self-view (such as hard anti-realists do) or by creating further definitional problems.

¹⁰ This problem, incidentally, will be looked at repeatedly in what follows in our study.

¹¹ Narveson, *ibid.*, note 35 on p. 18.

When we speak of the self in everyday conversation – whatever we may claim it to be if pushed to do so – we typically (unreflectively) mean the core or essence or ‘true’ nature of a person, not the person in her entirety. Would an element like one’s personality be included in this essential nucleus? It may or it may not be, but either way the self concept under discussion would change; how then might we ask the same question of a ‘person’? A person who has a personality included is just a person, and a person who does not have a personality included is the butt of jokes (surely some examples come to mind for the reader at this point). Similarly with bodily issues: the interpretation of the self involved will address these in one way or another, but to argue for either an embodied or a bodiless ‘person’ makes little sense, some segregation of self and person must be made before any physical claims could be considered. If, however, we want to speak of the self as being the ‘whole person’ then what we are really advancing is the absence of a specific ‘self’ altogether (which again is the hard anti-realist view and not the soft anti-realist view currently under discussion). If there is no primary and identifying centrality to the person then there is simply the person, warts and all as it were. This argument is an interesting one and it will be discussed below, but for the time being I simply wish to point out that if that is what is meant by the self under this account then we are no longer discussing a soft anti-realist self concept.

A number of proposals have been given to fill in the gap here left by the idea of the self as nonsubstantial but still experientially present, notably Daniel Dennett’s and similar narrative accounts of the self as the central character (or ‘center of narrative gravity’) in descriptions of what happens, is happening, and/or happened to a given brain-body aggregate.¹² It should be noted that by such solutions we still have a something foisted onto a nothing (a self without formative substance), but narrative answers do, perhaps

¹² Daniel C. Dennett, *Consciousness Explained* (New York: Little, Brown and Co., 1991).

unsurprisingly given Dennett's own neuroscience background, take into account the brain's parallel and distributed structure. What is that structure? Current research in the cognitive sciences maintains that the human brain has no 'central command center', but instead has a number of structures that are specific and locally processing, bound together in a network out of which the mind is thought to arise as an emergent property. (I note here that this is an important topic and will be returned to throughout.) Although we tend to naturally judge that experiential consciousness is a unified whole, it is thought by those who study the brain to be better described as consisting of a 'constellation' of differentiated and specialized consciousness systems whose products are integrated and interpreted by a cognitive module that evolved for that purpose.¹³ While it may be tempting at this juncture to go ahead and label that interpreting module as being what the self really is, even were that module to be definitively proven and its location – if it has one – discovered in the neural mass, its function would remain descriptive and not generative. To call this the self would be like calling a news report about an event the event itself. Towards this main issue of substance on the soft anti-realist view, therefore, 'central character' type answers are helpful but still do not seem to provide much beyond a kind of inner voice bandage placed over the core problem/question of 'what'. It seems that either a hard anti-realist stance must be taken or an account of contents stipulated.

Moreover, as Kristjánsson has pointed out, soft anti-realist positions, be they of the narrative or another variety, cannot differentiate between self-knowledge and self-deception:¹⁴ I may think that something happened to me only later to be told that that actually happened to my brother and I was misremembering what he had told me; before being informed of my error I was convinced that what happened to my brain-body was the

¹³ Michael S. Gazzaniga, *Who's In Charge?: Free Will and the Science of the Brain* (New York: Ecco Press, 2011).

¹⁴ Kristjánsson, *op. cit.*

plotline of the story in question, and its central character was me. After learning my mistake, what is the 'real' identity of the character in the erroneous section of my own personal narrative? While I held the false memory were my narrated self and my brother's narrated self somehow both the same, had his self entered my account or mine his? Where is the 'soft' anti-realist self traceable to? If nowhere, then there is no pinpointable central character, and if somewhere then in merging both characters have (admittedly temporarily, as it turns out) lost their distinctness as characters. Additionally, in a case such as this, where was my 'real' self in the duration of that lost time in which my memory has deceived me into narrating falsely? Who – and/or where – was the 'me' of the matters that really did occur? In such situations we cannot simply write off the false memory as an incident of being mistaken as in this line of thought the narrative (or string of owner-owned relationships) is, for all practical purposes, what the self is considered to be. If we simply say 'I was wrong' then the whole 'I' falls away. By skirting the line between the anti-realist and realist positions (taking the self to be something that cannot be definitively outlined but that is still in some manner present), this self concept is attempting to maintain a trackable self that is nevertheless nonsubstantive. Thus, to borrow other narrative imagery, instances of erroneous memories are akin to having a chapter from a different book inserted into the volume that one is currently reading (or writing, if the 'author of my own life' sense is being stressed). This objection is not a claim regarding lost time (e.g. blackouts, deep sleep, or other periods of seeming non-awareness or forgotten incidents); it is rather an assertion that these accounts cannot adequately address certain points related to matters of self-knowing and failing to know. In more technical terms, without a clear metaphysics epistemology suffers.

Finally, we may note that the soft anti-realist self, with its viewpoint of sealed off and atomized individuals, fails too to notice that no living creature can exist in a vacuum and

that each human (and nonhuman) self is highly contextual and circumstantially formed, based not only on current situation but also local culture, historical time period, geographic location, and a host of other details.¹⁵ The self cannot be defined in the absence of such because the self will never exist in the absence of such, nor will the self ever be fully free from outside influences that affect behavior and decision-making, a point we will come back to often in what follows. For the time being, however, we must move our attention to the other representative anti-realist account, one that has already been mentioned in the above commentary: that of the hard anti-realist self.

B. The hard anti-realist position

As with the soft anti-realist self concept, this subsection will first present the hard anti-realist position by allowing it to speak for itself, and only thereafter offer some brief criticisms. Once more generalities will be favored to depth as our goal here is simply to provide an overview of current self theories in sufficient breadth that the reader may be well placed for the following chapters, in which the analyses will be far more detailed and (hopefully) nuanced.

This account endeavors to rectify one of the central errors of its soft cousin: that of attempting to be a something with nothing underneath. The position is laid out in great intricacy by Derek Parfit in his *Reasons and Persons*,¹⁶ there called the reductionist view. To make his case for a total absence of self Parfit uses a series of imagined scenarios of

¹⁵ Heidi M. Ravven, *The Self Beyond Itself: An Alternative History of Ethics, the New Brain Sciences, and the Myth of Free Will* (New York: The New Press, 2013); see also the discussions on embeddedness and Edmund Husserl's and Martin Heidegger's important concepts of 'horizon' and 'world' (respectively) in the below, found in Chapter 2, Section 2.2.D and Section 2.5 ('Psychological issues and the self: Selves, personal identity, and whole persons', and 'Certain uncertainty, randomness, and limited choice'), and Chapter 3, Section 3.3.C and 3.3.D ('Phenomenologically-based accounts and related issues: Phenomenology 3: Going back to Husserl', and 'Phenomenologically-based accounts and related issues: Phenomenology 4: Heidegger on the self').

¹⁶ Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984).

varying degrees of likelihood (such as a false memory about a day in Venice or a teleportation to Mars) to arrive at and then expand on the conclusions he uses to define the hard anti-realist viewpoint: 1) we do not exist separately from brain/body, physical/mental events, 2) identity is not always determinate, 3) unity of consciousness and the unity of a whole life cannot be explained by claiming many ‘different experiences are had by the same person’ but must describe the relations between the experiences and the person’s brain, and can be fully described without claiming the ‘experiences are had by a person’, and 4) personal identity does not matter most and what does matter most is psychological connectedness and/or continuity with any cause for such.¹⁷ This final conclusion – ‘what does matter most’ – is labeled relation R, which Parfit defines succinctly as: ‘psychological connectedness and/or psychological continuity, with the right kind of cause’, adding ‘in an account of what matters, the right kind of cause could be any cause’.¹⁸

It may be helpful at this point to introduce an example of what Parfit means, and so we will look at one that he himself uses, namely, interplanetary teleportation. Imagine that you have been sent to Mars by your employer on an assignment, but due to the limitations of space travel the fastest way to get there is not to be physically transported but instead to use a teleportation device,¹⁹ a machine that does not actually move anything but rather reads it and reproduces it elsewhere. The way the teleporter works is that it scans the current state of every cell in your body and then fully reconstructs an exact replica at its sister location on Mars, eliminating your Earth body in the process.²⁰ In such a case all

¹⁷ Parfit also allows that physical continuity and similarity may be important too; *ibid.*, pp. 216-217.

¹⁸ *ibid.*, p. 262.

¹⁹ Parfit actually calls it a ‘teletransportation’ device; I have shortened it for simplicity’s sake.

²⁰ In his thought experiments Parfit considers both cases of the machine destroying your body on Earth in the process of replication, and that of your Earth body not being destroyed; in the two body scenario what is essential – relation R – continues for each as both the Earth and Mars bodies have psychological connectedness back to, and continuity with, the original source body. In this way the person has doubled but relation R has not split, rather it has continued in full in two different directions; this idea is picked up again below in the discussion on ‘branching cases’.

that you are left with in your brand new body on Mars – which is ‘brand new’ but still exactly like your pre-teleported body – is relation R, but this does not mean that ‘you’ (the self) are relation R because, again, Parfit stresses that there is nothing there of substance post-teleportation that has remained. There is no self, and we deceive ourselves into thinking that there is only because we typically have both relation R and physical continuity; what Parfit means to demonstrate is that there are scenarios (albeit at present fantastical ones) that challenge these presumptions yet still force us to admit that I am indeed my Mars ‘me’ without there being anything of ‘me’ on Mars at all.

Parfit finds this very liberating, proclaiming, ‘On my view, what fundamentally matters, in our concern about our own future, is the holding of relation R, with any cause. This would be what matters even when it does not coincide with personal identity.’²¹ On that latter, for example, instances where relation R does not coincide with personal identity would be like the one just discussed, a ‘branch-line case’ where your Earth body stops but your Mars body begins, or conversely such could be a ‘branching case’, where your Mars body begins even while your Earth body goes on (due to the teleporter’s successfully replicating your body on Mars without having to destroy your Earth body), thus making both bodies ‘you’; and since there is no self, both bodies *could* simultaneously be ‘you’ in the sense of the continuation of what Parfit thinks does really matter, the indispensable information that is relation R. As such the ‘you’ that has doubled must not be understood in the Cartesian or soul manner in which we have traditionally considered the self; Parfit is seeking to break us out of that mode of thought and is not claiming anything along the lines of Earth/Mars clones who would henceforth (post-creation) remain identical in every respect.

That something like this is imaginable potentially takes the sting out of death – Parfit states

²¹ Parfit, *ibid.*, p. 289.

that it does for him, anyway – and allows us, he maintains, to view our own mortality as just one more blip in the long stretch of the natural world’s cycle.²² There is nothing that is essentially me, so when I die nothing will be lost. Indeed, by this view, we may ‘die’ any number of times during our lives as relation R is dropped and taken up again.²³ Parfit tells us, ‘If we are Reductionists, we regard the rough subdivisions within lives as, in certain ways, like the divisions between lives. We may therefore come to treat alike two kinds of distribution: within lives, and between lives.’²⁴ In its full denial of any self the hard anti-realist account may be counterintuitive yet it still appears to be a clear and strongly argued self concept.

It is not, however, without its problems. To begin with, there is the matter of personal identity and Parfit’s claim that relation R matters more than it. I should signal at this point that since there is no self in hard anti-realist self-views those who argue for them tend to use personal identity as substitutable for or equal to what is called ‘the self’ in other accounts: the ‘I’ in ‘what I really am’. In considering what we think we are, that is, what our personal identities are, Parfit examines both physical and psychological criteria and declares each to actually be reductionist (i.e. agreeing with his own position). The physical case is that a person is the same person if enough of the brain, and not the whole body, has continued (without branching) between the past and the present; the psychological case is that a person is the same person if there is overlapping psychological connectedness that forms a psychological continuity between the past and present, also without having

²² *ibid.*

²³ For example, your early childhood years of which you have no recollection when shown photographs of the time, or even blacking out what happened after your staff party the night before. Such psychological schisms should not be confused with the types of starts and stops that we experience when going to sleep and waking up the next day; these are deep breaks but they are not unknown and most of us will have experienced them at some point. Although we lose relation R in these cases our physical bodies keep going and relation R is begun again, but because relation R has first been halted (the result of the break) we may, on Parfit’s account, think of the experience as actually having died, as having lost that which does really matter.

²⁴ Parfit, *ibid.*, pp. 333-334.

branched. These boil down to Parfit's view, he writes, because: 1) 'the fact of a person's identity over time just consists in the holding of certain more particular facts', and maybe also 2) 'these facts can be described without either presupposing the identity of this person, or explicitly claiming that the experiences in this person's life are had by this person, or even explicitly claiming that this person exists. These facts can be described in an *impersonal* way.'²⁵ We must therefore reject both of these conclusions to not accept as accurate a hard anti-realist self theory, Parfit states, and instead take personal identity as involving something beyond what the physical and psychological cases are propounding: a further fact of the soft anti-realist or realist kind.

These are well argued stances, as stated, and to begin our response to such reasoning we can grant Parfit the point that the physical and psychological arguments as he has stated them reduce to his reductionist position while still disagreeing that relation R ('psychological connectedness and/or psychological continuity, with the right kind of cause') has more value than personal identity. We can think this, moreover, even allowing the substitution of personal identity for self because what each of us takes our identity to be is singular and resonates with deep personal – and only personal, it cannot be transferred – meaning for us. This meaning by itself is arguably enough to constitute a 'something beyond', a 'further fact', an indescribable internal idea of the fullness of oneself that results from the lived experiences of one's life, a totaling of everything in, from, and about one in the moment to moment, an idea that most of us would be very hard pressed to put into words if requested to do so. We might say, 'I am such and such a person' or, 'I strive to be more patient', or 'I am a medical doctor', or any number of descriptors which are the types of things that Parfit seems to have in mind with his 'particular facts'. Yet would these statements, or a thousand of these statements, cover the entirety of a person, especially if

²⁵ *ibid.*, p. 210; emphases in the original.

considered from their own internal point of view? Would I ever really think that being a doctor or trying to be more patient or being a doctor *and* trying to be more patient (et cetera, et cetera) would constitute the whole of what I am? I doubt it, I think that for myself and for most anyone there is a deeper meaning to identity, and it is this meaning, this ineffable but essential notion, that I think connects a present person with their own future version in a way beyond how R is defined. This requires further analysis.

Parfit recognizes and tries to preempt the kind of uniqueness objections such as mine just given by the following: He concedes that when R is held uniquely (U) in a one-one form (i.e. nontransferable (and nonbranching in teleportation or other imaginary cases)), then personal identity (PI) is equal to relation R plus uniqueness, thus: $PI = R + U$. Parfit next reduces U by stating that if I am R related to a person ‘the presence or absence of U makes no difference to the intrinsic nature of my relation to this person’, and hence it adds very little to R, although Parfit admits that uniqueness ‘can be plausibly claimed to make a small difference’. Yet this ‘small difference’, Parfit then asserts, ‘would be much less than the intrinsic value of R. The value of PI is much less than the value that R would have in the absence of PI, when U fails to hold.’²⁶ I find this to be technical and somewhat slippery writing that calls for a careful examination of just what is being put forward (and, considering the first person contents of a concept like my own uniqueness objection, how could uniqueness ever fail to hold?).

On Parfit’s own terms, using his variables as he sees them, let us attempt a dissection of this stance. If we only had R in the absence of PI what would we have? This would not be a case of ‘dying’ within a lifetime (psychological schism), for in such cases relation R itself is lost; here now we only have relation R continuing and everything else disappearing. This

²⁶ *ibid.*, p. 263.

must be something else. What is really at stake is the elimination of U – this is after all a response to the uniqueness objection – and Parfit says that the most forthright example we can imagine of such a loss of uniqueness would be a division case similar to the one discussed earlier where an Earth body continues even while a Mars body begins, but instead of the situation where after the original Earth body is destroyed only a Mars body is made (the scanning and reproduction process involves the body's destruction), now both a new Earth body and a new Mars body are created by the teleportation machine: two duplicate bodies from one source. Thus my parenthetically placed question above is answered: Parfit thinks uniqueness could fail to hold in a (by present technology) fantastical scenario like that provided by transplanetary replication. In such a case each of the new bodies retains relation R, but it is not held uniquely, and whatever personal identity the two have must either stem entirely from R (as that is all that remains) or not yet be formed. If their personal identity(ies) results only from the holding of R, Parfit argues, then we must assume that both bodies will have the same personal identity (and therefore no uniqueness), and if in both it is not yet formed then the bodies would only have R and nothing else: no personal identity and no uniqueness. Either way, Parfit concludes that R would indeed hold the greater value than the other elements.

Let us play along with this science fiction game and ask: Would this result really be what happens? If I imagine myself entering a machine here on Earth one moment and the next moment exiting a similar machine on Mars (or the same machine on Earth) I cannot see how that would eliminate the holding of my personal identity (including what I have suggested is the meaningful, meaning-making 'further fact') and leave me with *only* a psychological connectedness and/or continuity. I cannot see this because I cannot see how uniqueness (and hence personal identity on Parfit's own equation ($PI = R + U$)) would ever be definitively lost; my parenthetically placed question is still with me. Why?

Experientially speaking, as far as I know I am still the same old me, even if I were told beforehand that all of the cells in my body would be replicated and that part of that process meant the destruction of my old cells.²⁷ I would still be receiving all of my sensory input in the same ways, still see the same face in the mirror, still look down at the same hands and body below my neck and still carry the same information stored in my brain as before the procedure. What has happened has been instantaneous, and even if it were not instantaneous the end result of an exact replication of an original would still be the same, there would simply be a break of X amount of time in between (and presumably while this might indicate a potential loss in some manner, experientially I would again only step into a machine and then step out, and for me no ‘personal time’ would have been lost²⁸). Moreover, stepping out of that machine in one location or the other would immediately begin the establishment of a uniquely held personal identity even if we grant that in the fraction of a second when both bodies had only just been replicated and neither had blinked an eye to see where they now were it was not technically held uniquely according to how Parfit interprets uniqueness using the features of his scene setting (although each would still *consider their own* to be unique – this is an important point and that is why I keep coming back to it). Nevertheless, even giving Parfit this fraction of a second each version of the original ‘I’ would still maintain their internally held personal identity, such would yet mean a great deal to each of them, and it would furthermore increasingly diverge from the other-planetary double’s personal identity with every passing second as their lives moved on and fresh experiences inevitably added up.²⁹ There seems to be no way of actually ever losing one’s sense of uniqueness, and not only one’s sense of it but

²⁷ Remember that in this thought experiment the machine creates *both* a new Earth and a new Mars body; see the summary of Parfit’s description in the previous paragraph.

²⁸ The experiential feel of time will be important in the final chapter of our study.

²⁹ Relation R would also diverge as time went on, of course, eventually leaving each double with potentially no connection to their original singular being (as in the case of the non-recollection of childhood events). The personhood of each is in question here, and that issue is debatable (note though that in both the physical and psychological criteria there is a ‘no branching’ rule), but the point that I am trying to make is not about personhood, only that relation R does not matter more than personal identity.

definitionally the fact of it too. My parenthetically placed question remains.

Readers sympathetic to anti-realist accounts may however wish to object that that fraction of a second before either body blinked is all that is needed to establish that relation R has more value than personal identity. In that thought notice, though, that we have left something important out of all this, namely that heretofore we have only considered one temporal direction of the relation R and personal identity question: that of the past. We have focused on keeping the relation R link between the present replicated (teleported) body and its pre-replication source body, and nothing more.³⁰ Yet as Carol Rovane has pointed out, we are bound to our futures through our forward-looking attitudes that are rooted in first person psychological states (such as memory, anticipation, and intention),³¹ and such attitudes must be based on personal identity. Yet again we find that notion of one's own (meaningful) notion of 'I'. Without a sense of identity we would have no reason even to contemplate our future: there would be no 'me' in my future if I failed to hold a referencing personal identity. Although we may lose our memory of the past, such as in cases of severe amnesia, we never lose that forward-looking sense with which we are so familiar. We might even immediately forget each moment as it passes if we suffer from a peculiar form of brain damage or disease, yet in each now we would yet anticipate and intend. Once more, we can afford to be generous in our argument and consider that if we do agree with Parfit's relation R conclusions we may even think that we lose it at times and 'die' within a lifetime (as above), but if we are aware we will always maintain an internal and personal future-oriented perspective, and that must fundamentally include a minimum

³⁰ While it is true that we have been imagining these scenarios taking place in the future what has been the focus of everything has been simple continuation, that connection between the 'now' and the 'then'.

³¹ Carol Rovane, 'Branching Self-Consciousness', *The Philosophical Review*, 99:3 (1990), 355-395. Although Rovane's article is meant to defend a reductionist view by explaining how it can accommodate future interests (by arguing that since we anticipate a psychological connection with our future self(ves) we identify with them as a person particularly capable of carrying out our projects, and that indeed such 'identification' really comes down to the anticipation of relation R obtaining), what is important for our purposes here is its (rightful, I think) emphasis on the future.

of personal identity, an association of a continuing ‘me’ somehow connected to a ‘my’ body (which may or may not be replicated, and if replicated yet remains an exact copy of the original – all that changes none of our conclusions), if nothing more. There is no way to bend these playful pictures of replication where relation R comes out on top.

We now have a claim (claims, really) contrary to Parfit’s assertions that relation R matters more than personal identity and that personal identity does not matter most (recall his initial conclusion 4) listed above: personal identity does not matter most and what does matter most is psychological connectedness and/or continuity with any cause for such), which we can add to the more commonsense objections to the hard anti-realist position of which Parfit’s account is representative. Foremost amongst these latter rebuttals, as the preceding has endeavored to show, is that by this view things simply do not work. Kristjánsson writes that hard anti-realists ‘take pot shots at the notion of truth as correspondence with reality. There is a basic difficulty with rejecting this notion: Almost all human actions, communications, interactions and investigations seem to presuppose its truth.’³² (That Parfit’s arguments rely on scenarios that themselves have no correspondence with reality may be instructive in this regard, but their usefulness as analytic tools could belie that.) Furthermore, without at least a sense of persisting internal sameness that lasts over time, moral responsibility becomes increasingly difficult to pin down.³³ If we can indeed treat subdivisions within lives as being like different lives in some ways then we may find ourselves embarking on a path that results in entirely undesired reforms to our justice systems. (E.g. Is Frank in his current life responsible for the crime the entity we used to call Frank – an entity who looks just like our Frank now but who experienced such a subdivision – committed prior to the subdivision? What if he is

³² Kristjánsson, *op. cit.*, p. 38.

³³ *ibid.*

found not to be?³⁴) In thinking along these lines we may find ourselves agreeing with Kristjánsson that on the whole the hard anti-realist viewpoint, ‘makes travesty of everyday moral experience’.³⁵ It is also, we might add, unclear how the ubiquitous situational and contextual pressures that we all face would affect one on this account, or even if they would be thought to affect one at all. Would such need to be considered only in their influences on relation R? Or would personal identity now come into play for Parfit? If so, how? Why here? There might be answers to all these questions but we are left to fill them in for ourselves. Finally, as we noted above, emergent properties can exist physiologically: the sense of a unified conscious mind may be a primary example of such if current cognitive models are correct. Could the self not therefore also be some type of emergent property of natural biological functioning, even if it cannot be pinned down in a hard sense? Might that be the basis for the internally held meaningful idea of ‘me’ that I outlined as an objection to Parfit’s account? I think that indeed this question lies at the very core of any workable realist conception of the self, and it will be returned to in Chapter 2 where our proposed alternative self-view will be fully fleshed out, and then considered again from a metaphysical standpoint in Chapter 6. First though the next two subsections will complete our overview of currently existing representative self accounts by examining the soft realist and contextualized soft realist positions.

C. The soft realist position

This subsection will match its predecessors in beginning with a presentation of the position as its representative account describes it before then offering some criticisms, expansions and explanations, and comments on how the view might be used. Perhaps more so than the previous two self concepts explored the soft realist position may at first appear to be

³⁴ Other undesired judicial changes stemming from different reasons are also a possible outcome of the contextualized soft realist account discussed below.

³⁵ Kristjánsson, *ibid.*, p. 46.

deeply confused, but I believe it is well worth the effort to explore it, that there is much of value within it, and indeed this and its related investigations will become the focus of the remainder of our study. In the below, however, we seek only to offer generalities and expansive brushstrokes with our inquiring pen, and then thereafter with what follows from here in mind will we be able to find our way through the more detailed and precise analyses that will follow in Chapters 2 through 6.

What may perhaps strike many readers as being most remarkable about this account of the self is that it is realist: that it claims that there is in fact a substantive nature to selfhood without making the further Cartesian ego or soul claims that we tend to associate with the realist stance. Kristjánsson begins his case for this self-view by admitting that while there is no consensus on the definitive way to interpret Book Two of Hume's *A Treatise of Human Nature*, he thinks that Hume 'seems to be arguing that, whereas the self as a succession of related ideas and impressions cannot be a direct object for the understanding, the self of whose moral actions each of us is intimately conscious can be a direct object for our emotions', making the self's realism, its actuality and its permanence, consisting of internal emotional activity that is further dependent upon external reinforcement from others via social interactions and societal rules and conventions regarding the emotions.³⁶

Although it is not entirely clear exactly what Kristjánsson is driving at in this his main work on the subject of the self (which splits its focus between self theory and educational/therapeutic applications), he appears to be using Hume's idea as a means of moving away from the Lockean concept of a self-understanding based on the continuity of consciousness and conscious access to memory (or at least the possibility of being able to consciously access memory (thus accounting for breaks of forgetfulness, sleep, and the

³⁶ *ibid.*, pp. 47-48.

like)),³⁷ and instead moving towards his own interpretation of Hume wherein self understanding is more properly based on one's shifting emotions, subject to both personal reflection and societal support and/or censure. The self here in Kristjánsson's position is taken to be each person's moral being, the so-called day-to-day psychological unit of reference that one goes by, akin to 'the voter' or 'the citizen' or 'the taxpayer'; it is seeing oneself from an affective and morally related point of view.³⁸ Kristjánsson summarizes his soft realist self as being composed of three sets of self-related emotions: 1) Self-constituting emotions: those that define us, our 'core commitments, traits, aspirations or ideals', 2) Self-comparative emotions: those that take the self as 'an indirect object' or 'a reference point' for 'comparison with a baseline of expectations', and 3) Self-conscious emotions: those that are in the self they are about, that take the self as 'their direct attentional and intentional object'.³⁹ The self on this account is that which we maintain as a personal inner point of reference, and that more or less matches what we mean when we speak of ourselves; it is the culmination of the creature performing the actions, thinking the thoughts, and having the feelings that we internally associate with those actions, thoughts, and feelings, and that others associate with them too. There is no mental 'pure ego' here that takes up residence in the physical body (the 'psychological unit' mentioned might best be understood almost as a side effect or default result instead of as the core essence that a 'pure ego' would be), rather on this self theory the physical body, along with a unique sense of personal identity, is all that the self is, and a part of that body's normal functioning is to have the emotion-based ongoing characteristics that make up Kristjánsson's first emotive set and the self-referencing features that make up his second and third sets.

³⁷ See Chapter 27, especially sections 11-12, 'Of Identity and Diversity' in John Locke, *An Essay Concerning Human Understanding*, 2nd edn (1689/1690). Available on Project Gutenberg: <<http://www.gutenberg.org/cache/epub/10615/pg10615.html>>. Notice that here and in Kristjánsson (amongst many others) 'consciousness' is used synonymously with 'awareness'; in what follows we will be drawing a line between those terms and concepts, and that differentiation will, I think, prove important.

³⁸ Kristjánsson, *op. cit.*

³⁹ *ibid.*, pp. 75-77.

This view of the self, like the narrative variety of the soft anti-realist position, also matches with the multivariate and highly specialized modular functioning of our physical brains,⁴⁰ but unlike the narrative account Kristjánsson's soft realist self posits a substantive object (the 'day-to-day psychological unit') that can allegedly differentiate between self-knowledge and self-deception, even if it cannot be said to exist physically. Our own self concept as a moral being that is subject to the social, moreover, means that for each of us that conception can correspond to reality or fail to correspond, it can be judged objectively through the lens of its quality as other-dependent in the public realm. Daily interactions will either reinforce the way in which we see ourselves as being or they will show us where we have been in error.

Of the accounts of the self examined so far, this one appears to be the best situated initially in terms of being both theoretically sound and neurologically accurate (as regards functioning), yet it contains many philosophical problems that will need to be dealt with. As mentioned at the outset, I think that this position can be built into something that is usable and even advantageous, but that a great deal of work will need to be done on it before we can get to that point. The bulk of that labor will be conducted in the next chapter (supplemented significantly by further efforts made regarding the possible nature of substance as it is presented here (done in Chapter 6 via an alternative form of realism)), but in this subsection we can begin to sketch out how this self concept might be understood and expanded into something of more clearly recognizable robustness.

The first aspect to preliminarily wrestle with is how Kristjánsson's psychological unit of reference can be viewed as a realist and substantive self without it being something along

⁴⁰ Gazzaniga, *op. cit.*

the lines of a Cartesian ego. It must be remembered that the everyday internal item of reference being discussed is based on emotional subject matter, and as such it is always shifting, permanent in its existence (continuance) but not in its moment-by-moment content (the metaphor of a river readily presents itself at this point). Such a view is not one that takes the self and whole person relationship to be like a pilot steering a ship, but instead is something more like an organic digital device that is constantly updating (and that usually runs without being directed; in the below we will also consider in what ways the software/hardware analogy of the self is likely misplaced). The comparison to a digital device is admittedly stretched, and for personal reasons I dislike it and hesitate to employ it, but what I want to stress by it is that from this imagery we should take the notions of fluidity, mobility, and automatic renewal – prompted both internally and externally – that comes with the analogy while at the same time avoiding the type of static and ‘pure’ pictures that can be conjured when a computer or robotic machine is invoked. I think that we may consider Kristjánsson’s central idea to indicate that both how we understand ourselves when we purposefully think about ourselves, and how we understand our ongoing traits (of genetic and experiential and/or formative origin) when we are not thinking about ourselves (that is, how we are most of the time) together form that thing we mean when we say ‘my self’. Put another way, when we reflect on what we look like when viewing ourselves in a mirror (that is, when not actually physically looking), we have a certain outline that tends to come to mind and that will be largely accurate but perhaps not capture all of the details of our appearance at that moment. The soft realist self taken as our ‘day-to-day psychological unit’ might be like this: always with us in the background of our minds and mostly accessible when called upon, but still needing outside (social) help for complete accuracy.

Another problem that this account exhibits has to do with how Kristjánsson handles agency

and decision-making in his soft realist self. In a discussion of the ‘gappiness problem’ taken from research findings in a number of psychological studies⁴¹ which indicate that moral reasoning either fails to motivate moral action or does so only slightly, Kristjánsson refers to his ‘unified moral self of rationally grounded emotion’ as a means of repairing the disconnect – or rather as showing that there actually is no disconnect and that the root of the problem of failed moral action lies elsewhere.⁴² He does this by arguing that his soft realist self demonstrates that there is no difference between the ‘moral-self’ (having moral concerns as part of one’s identity⁴³) and having moral emotions, thus motivating moral action, because those emotions are the foundations of the self. Kristjánsson writes that moral emotion can join together moral cognition and moral action through the training, in an Aristotelean sense, of a (soft realist) self to be a moral self. This possibility is grounded in the theoretical distinction between episodic emotions and dispositional emotions (an idea we will return to through another researcher in the following chapter), with the moral (soft realist) self having its basis in the latter: the purported dispositional emotions. By acting out of that foundation any given emotional reaction will reveal the internalization and integration of ‘a certain emotional disposition into his or her moral self.’⁴⁴ In other words, Kristjánsson’s soft realist self-view has the self arising from the (trained or ingrained) emotional dispositions and therefore any (short-lived) episodic emotions will stem from that source and thereby express the content of the self in question. Instances out of traits, if you will. It appears that what Kristjánsson is getting at in his discussion is that a person who has willingly implanted within themselves certain character features has consequently established preferred automatic reactions, and these will naturally manifest in their respectively generated moments. A truly moral self, therefore, one whose internal

⁴¹ Augusto Blasi, ‘Bridging moral cognition and moral action: A critical review of the literature’, *Psychological Bulletin*, 88:1 (1980), 1-45.

⁴² Kristjánsson, *op. cit.*, p. 97.

⁴³ This was Blasi’s suggestion for how to solve the gappiness problem; Blasi, *op. cit.*

⁴⁴ Kristjánsson, *op. cit.*, p. 94.

image as a moral person matches with how that person is viewed by others in the world, will exhibit stably moral behavior arising out of their structurally secure emotive reflexes.

This reasoning, however, strikes me as confusing an emotional response with an intuitive response (and again, this criticism will be returned to below), and it moreover fails to note the evolutionary grounding that our emotions have, giving rationality a more central role than I think it is capable of taking. Kristjánsson's account has it that an active and reflective decision is required to make moral concerns part of one's self identity and that, again, the baseline of any moral self is 'rationally grounded emotion',⁴⁵ yet there is a large amount of empirical research that appears to take us towards conclusions very far from this. Jonathan Haidt has taken the lead in arguing that an ever-increasing pool of cognitive and psychological data has demonstrated that most decisions are in fact made preconsciously⁴⁶ and not reflectively, that they are based on intuitions that have been honed by natural pressures over millennia, and that only after a decision has been made and acted upon (or thought or feeling generated) does rationality step in and provide an internal reason for the action, thought, and/or feeling.⁴⁷

In fairness, it might be that Kristjánsson in his writing was not concerned with decision-making as such, but rather meant only that one needs to determine (in a virtue

⁴⁵ *ibid.*

⁴⁶ For reasons that will be explained in more detail in the following chapters I prefer the term 'preconscious' to 'unconscious' or 'subconscious' as I think the nuance provided more accurately fits with current models of how our brains work, and moreover indicates a closer connection between non-aware and aware workings. While it is true that in some ways what I wish to express by 'preconscious' is closer to 'nonaware conscious systems' or, perhaps better, 'mental but not available to awareness', those terms are cumbersome and carry varying amounts of linguistic baggage. To me the important point is that in common usage 'unconscious' versus 'conscious' has an underlying intimation of a break or dividing line that I do not see as being there – or at least not there as sharply as the terms imply – but I may be in the minority in thinking that. At any rate, my argument will be given in full in the below.

⁴⁷ Jonathan Haidt, 'The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment', *Psychological Review*, 108:4 (2001), 814-834; see also Jonathan Haidt, *The Righteous Mind: Why Good People Are Divided by Politics and Religion* (New York: Pantheon Books, 2012). Each source summarizes the evidence and presents Haidt's Social Intuitionist model in detail.

ethics sense, or more broadly Aristotelean training sense) that one wishes to be a morally concerned person, but if that is all he is indicating then the argument strikes me as a somewhat shallow one. After all, few of us completely avoid at least attempting to be morally concerned in one way or another. Is it not the case that we are by default – as social creatures – attentive to our relations with others and hence morality (of some type) automatically comes into play? Is it not rather that the choice to be consistently *immoral* in any and all dealings is the one that really requires some effort? Let us therefore assume that Kristjánsson's model of 'rationally grounded emotion' does make room for behavioral determinations at the everyday level and continue our exposition of what the cognitive sciences have indicated in this area.

Essentially, the research has revealed that what the process of our preconscious decision-making seems to entail is that we encounter a situation or receive – even simply perceive – a stimuli which our automatic and non-aware conscious systems (i.e. those never entering into purposive awareness, those parts of consciousness that remain 'preconscious' in the sense being used here, traditionally labeled as the 'subconscious') interpret and then react to, generating a physical behavior or mental stance: disgust is a clear example of this. After such has been produced our consciousness (i.e. with awareness) takes note of said behavior/stance, and then (often) seeks to fill in the apparent blank of why X happened by providing itself (intentionally thinking) with a reason for the occurrence of the action or judgment that has just been actively noticed (e.g. 'That painting reminds me of dog feces, how repulsive!'). These foundational intuitive rules that drive our decisions seem to be both *prima facie* and reflexive, biologically useful for their efficiency even if not always correct.⁴⁸ Michael Gazzaniga summarizes the brain's functioning in this

⁴⁸ Magda Osman and Ruth Stavy, 'Development of intuitive rules: Evaluating the application of the dual-system framework to understanding children's intuitive reasoning', *Psychonomic Bulletin & Review*, 13:6 (2006), 935-953.

way: ‘Many moral intuitions are rapid automatic judgments of behavior associated with strong feelings of rightness or appropriateness... not usually arrived at by a deliberate conscious evaluative process that has been influenced by reason’.⁴⁹

Disgust is a good illustration of this because it indicates how commingled our emotions and intuitions often are. They are not, however, identical, nor do they serve the same narrow purpose even if both do inform an alike general determination. To speak of ranks or degrees between emotions and intuitions is perhaps not helpful, but these functions are certainly interactive at a very fundamental level of our biological operating. Evolutionarily speaking, emotions are for the promotion of rewarding behaviors and the prohibition of non-rewarding (often dangerous) behaviors, whereas intuitions are for the rapid and effective decision-making to facilitate such promotion or prohibition. We feel disgust and quickly pull back, successfully avoiding the offending object without any need for cumbersome calculated thought.⁵⁰ Both of these aspects will be analyzed further in Chapter 2, for now the distinction I aim to make between the two as the issue relates to Kristjánsson’s representative self-view will hopefully be clear enough for present purposes.

All of this is not to say that we never think, nor that we do not rationally decide matters. Our choices can be and are swayed by reason, particularly when coming from others and thereby having a social affect on us,⁵¹ and our intuitive evaluations can also change during the course of our lives, but for the most part by the time we have gotten to the point where we are rationally deciding what to do we are not coolly considering our now-for-now and

⁴⁹ Gazzaniga, *op. cit.*, p. 166; here once more the standard usage of ‘conscious’ meaning ‘with awareness’; as noted above, my reasons for distinguishing between consciousness and awareness will be included below.

⁵⁰ Both Haidt sources (2001, 2012, *op. cit.*) are useful guides to the issues under discussion; see also Joshua Greene, *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them* (New York: Penguin Press, 2013).

⁵¹ Haidt 2001, *ibid.*

now-for-then preferences and then weighing them against logic and the facts,⁵² we are instead attempting to explain to ourselves the decisions that our preconscious minds have already come to and put into play. There are of course certain types of decisions for which rational methods are clearly called for ('What kind of house should I buy?'), but the empirical data gathered thus far indicates that even those will demonstrate the unavoidable intuitive reactions that influence and underpin all subsequent thinking ('Oh my, but I do like the look of that kitchen.'). Understanding this is important because any decision-making model that ignores this evidence will be largely inapplicable in the real world of practicalities. If a self concept is to be profitable (and not simply intellectually intriguing) it must take such into account.

It will be remembered that Kristjánsson's view does not claim to be a fully rational one, and in that it is nearer to the mark indicated by what we have been discussing than some others. Kristjánsson's account, after all, is based on emotion and seeks to promote individuals who will thoughtfully choose what kind of emotional dispositions they wish to have and will then set about inculcating them through a preferred or designated method(s). Yet this model nevertheless does still appear to be confused in the role it gives to the emotions (rather than intuitions), and in the weight it places on rationality with respect to decision-making. Although in some ways Kristjánsson's 'rationally grounded emotion' is similar to Haidt's work regarding the part intuitions play in decisions and behavior, it is in the reasons Kristjánsson gives for his rejection of the Haidtean view that we can see the main fault in his own.

Kristjánsson dismisses Haidt's and similar systems, as well as the opposing rationality-centered models, as being two-tiered (emotion on one level (recalling that

⁵² R. M. Hare, *Moral Thinking: Its Levels, Method, and Point* (Oxford: Oxford University Press, 1981).

Kristjánsson basically treats ‘emotion’ as both emotion and intuition in his usage of the term) and reason on another), where ‘emotion is distinct from – if complementary to – reason’,⁵³ and thus as not having placed one within the other, one as the other. Yet if the research revealing the timing of our decision-making (intuitions first, reasoning later) is correct, and at this point it certainly seems to be, then perhaps Kristjánsson’s order should be reversed, giving us a soft realist self of *intuitively grounded reason*. To do this we would need to replace Kristjánsson’s ‘emotional dispositions’ with intuitions, but as it appears that is what he is actually referring to in his discussion of dispositions (i.e. set modes that effortlessly yield reactions, including perspectives and/or judgments) such a substitution seems reasonable enough. That would, however, still leave Kristjánsson’s concern with the emotional (read: intuitive) training of a moral self intact, a point with important behavioral ramifications, for by the same cognitive research results it does also seem that we are capable of both honing existing intuitions and training ourselves to have new ones. Although intuitions are automatically generated decisions, determinations, evaluations, and/or inclinations made in reply to experiential stimuli and are further informed by emotional responses, they are not wholly inaccessible (i.e. not purely and fully genetically determined or the like) and we can equip ourselves with new or different initial intuitive reactions through both self-reflective and social input sources, a point that Haidt’s research highlights. This is where Kristjánsson’s conception shows merit, I think, as it emphasizes that to become the type of moral self that one wishes to be one needs first to examine and elect: to reflect rationally and arrive at a conclusion. If we do not want to, say, lash out at someone when they bump into us on the street if we find ourselves to be, or are told that we are, naturally inclined to do so, we first take on that commitment mentally and then actively seek not to engage in such behavior (or to engage in other forms) during our daily activities. We keep a watchful eye on ourselves and our natural (preconscious) reactions

⁵³ Kristjánsson, *op. cit.*, p. 98.

when people do bump into us, noting our emotions and behavior at the time and focusing on how we would rather proceed instead. Meditation programs and similar disciplinary methods can be very effective in such efforts at the re-training of one's reflexive behaviors, and something of that nature may be what Kristjánsson has in mind by emphasizing an emotion born of reason; his account simply appears to have mistaken emotions for intuitions and reversed the order of the foundational system, despite having the correct sequencing of steps for later affecting that system (using rationality to decide on how one wants to be).

There are still other problems that remain with the soft realist self concept as Kristjánsson has it, and these will need to be worked out, but those discussions can wait for the full explanation of the alternative self-view that I wish to propose in Chapter 2. Prior to that though, we will scrutinize the contextualized soft realist position as it also has beneficial points to offer, even if it too arguably cannot be adopted wholesale. For now, however, a final word on intuitions.

Considering all of the above, the question that most presses itself on us is this: Just what is the strongly intuitive and almost wholly automatic nature of our decision-making telling us? What are we to make philosophically of the things this psychological and neuroscientific research has revealed about the kind of creatures that we are with the type of natural functioning that we have? At once it appears to point to the limits of rational thought and reason-based argumentation, yet at the same time it also seems to indicate the potentially important way in which both can be used: for the furnishing of chosen intuitive reactions in those areas where control can be exerted. At present we do not know the limits of which intuitive reactions can be shaped and which cannot, but it strikes me as being plausible to speculate that those intuitions most closely associated with essential biological

needs will be less likely to be changeable to a consequential degree, and perhaps not at all. Other intuitive reactions, it would follow (if the above is correct), might be significantly malleable. What would be included in this latter category? Certainly much of what we do and feel that is under the current auspices of culture. What does this, in turn, suggest? That if our societies, with their natural and (mostly? internally partially planned?) unplanned evolution have the influencing power over many of our core intuitions that they appear to (an appraisal commonly arrived at via observation of variation between cultural groups and similarities within them), might not a more philosophical approach allow for a greater contemplative intuitive set? A person apt to reflect instead of react? In his Aristotelean concerns Kristjánsson at least seems to think so (though rather than an ‘intuitive set’ he would call them ‘emotional dispositions’), and I think many of us would agree on the desirability of such. At this point, however, I cannot offer more than these very general musings without taking us too far afield from our review of existing representative self concepts. We can though perhaps summarize these thoughts with the proposition that philosophy can be favorable to helping ascertain the type, extent, and manner of intuitions that we ought to desire and seek to possess. Philosophical argumentation could also be used to help convince us of what type of intuitive training programs we should pursue and for what ends (as well as for concerns relating to need, implementation, et cetera); such may grant some noteworthy benefits.

D. The contextualized soft realist position

The final self account that we will examine will once more be presented in its full before embarking on any critical comments and noting what the more applicable points to the forthcoming alternative self-view might be. This last self concept is similar to Kristjánsson’s in being a soft realist position, but is more inclusive of scientific research into cognitive functioning in its theoretical framework and far more radical in its

conclusions. This is the contextualized soft realist position of Heidi Ravven, applying a label that neither Ravven nor Kristjánsson has given to her work but one that I hope will not be misleading.

As with Kristjánsson's, this account posits a substantive self that is grounded in emotion, but it stretches that foundation to biological layers far below where Kristjánsson has them. Ravven sources rather heavily the neuroscientist Jaak Panksepp's work which argues that the self emerged evolutionarily as an affective system to facilitate survival, that it stems from very early sections of the mammalian brain, and that it is something which is shared across all species of mammal.⁵⁴ As such, this neurologically defined self is our (and not only 'our' but also nonhuman mammals') 'point of view of survival' and 'first emerges in the precognitive ability of most organisms to operate from an ego-centric point of view.'⁵⁵ Although this self also claims an emotional basis, it does not structure itself in a way like Kristjánsson's purportedly Humean reflective self is composed – with its three carefully sorted sets of self-related emotions –; rather, the emotional content under consideration here is less detailed in subject matter, akin more to a primal urgency, a voice from deep within a creature's subconscious (preconscious, on our terms) that guides and directs.⁵⁶

Ravven takes the claims of this biochemically oriented mammalian commonality and expands them socially outwards in the case of human beings, revealing a theory that is highly contextualized and not only substantive but situational, and mostly deterministic. She stresses the embedded nature of the composition of the self, stating that it is

⁵⁴ Jaak Panksepp, 'The Basic Emotional Circuits of Mammalian Brains: Do Animals Have Affective Lives?', *Neuroscience and Biobehavioral Reviews*, 35:9 (2011), doi10.1016/j.neubiorev.2011.08.003; in Ravven, *op. cit.*

⁵⁵ Stephen T. Asma and Thomas Greif, (2012), 'Affective Neuroscience and the Philosophy of Self', *Journal of Consciousness Studies*, 19 (3-4): 6-48; in Ravven, *op. cit.*

⁵⁶ Parallels are drawn between the biological view here and Spinoza's use of *conatus* (in the 'will to live' sense). Although Ravven does refer to Spinoza's *Ethics* (1677) throughout her work, she bases her self concept on her interpretation of contemporary neuroscience and not on Spinoza's *conatus* per se.

constructed by its current relation to another: that is, that we all have multiple selves each of which corresponds to a significant relationship and is partially formed by that very relationship and its object (other) through the carrying over of the sense of self involved. Yet despite this position's stance as a soft realist conception, none of us are thought by it to have a singular 'me'; instead, Ravven states, 'the feeling of self is a mental capacity that can be projected inward or even outward onto the world...we make parts of the world feel like self, and we fill our feeling of self with our engagements in the world.'⁵⁷ This notion of a malleable, distributed self that twists and turns with the contextual forces around it and that re-forms – re-invents, re-creates – itself with each new relation also leads Ravven to argue that our actions and fates 'are determined...by who our parents were, what world and situation we were born into, and who we became as a result of our early experience, our genetic inheritance, and on and on'.⁵⁸ Accordingly, Ravven contends that the agency we assign ourselves mistakenly infers causal ownership of actions when in fact there stretches behind each act a multitudinous number of causes and conditions, and that we ignore contributing factors and falsely imagine our behavior to spring from an unbiased free will.⁵⁹ Despite these assertions, Ravven does try to assign moral responsibility for individual activity by following Jonathan Lear's lead in his analysis of Oedipus: that, despite all that may have been ordained by outside forces, what matters in the end is that it was done by me.⁶⁰ Ravven writes that, 'This given "me" is that by which I am constituted. It is the "me" I find, and I resign myself to accepting it. In so doing, Lear says, I become transformed from being passively acted upon into a morally responsible agent.'⁶¹

⁵⁷ Ravven, *op. cit.*, p. 372.

⁵⁸ *ibid.*, p. 348.

⁵⁹ *ibid.*

⁶⁰ John Martin Fischer has taken a somewhat similar approach in his attempt to build a theory that lies between hard determinism and full free will; for an introduction to his ideas interested readers may wish to see *Deep Control: Essays on Free Will and Value* (Oxford: Oxford University Press, 2012).

⁶¹ Ravven, *op. cit.*, pp. 348-349; Jonathan Lear, *Love and Its Place in Nature: A Philosophical Interpretation of Freudian Psychoanalysis* (New Haven, CT: Yale University Press, 1998).

There is much that I think is good in the socially-minded and contextualized aspect that informs this position of a soft realist self, taking into account as it does the lessons learned of the overwhelming influence of group and setting from such famous psychological experiments as Philip Zimbardo's 1971 Stanford Prison Experiment and the electric shock experiments studying the tendency for obedience to authority figures by Stanley Milgram, both of which Ravven refers to on numerous occasions. Moreover, factors such as one's upbringing, historical time and place, socioeconomic background, and the capriciousness of the genetic lottery all certainly play large parts in our lives. However, to assign full determinism is a step that many will see as going much too far, and Ravven does more or less do so, if not in so many words. Her version of agency and moral responsibility, for instance, amounts to little more than an admission along the lines of: 'This thing that I did was generated by uncountable and interrelated background causes over which I had no control, culminating in the performance of the act by my physical body, but I will accept the consequences of the action anyway.' Very few of us would consent to taking on responsibility in this way if we held such a view (aside, perhaps, from an attitude of simply (possibly resignedly) surrendering to whatever happens to one; but that in essence seems to be what Ravven is getting at, unless I misread her), particularly in legal contexts. Moreover, if this account is true then our legal systems themselves would need an overhaul of a proportion that is difficult to even imagine. Gazzaniga, in his consideration of social influences and the brain's highly programmed and often unreachable workings that occur automatically and preconsciously (without ever entering awareness), still concludes that 'ultimately responsibility is a contract between two people rather than a property of a brain, and determinism has no meaning in this context... Brakes can be put on unconscious intentions', adding that in the end 'we have to look at the whole picture, a brain in the midst of and interacting with other brains, not just one brain in isolation.'⁶² If one concurs

⁶² Gazzaniga, *op. cit.*, p. 215; note again that Gazzaniga is using 'unconscious' in its standard way to equate

with Gazzaniga's analysis, and it appears reasonable enough to at least tentatively do so (tentative given the rapid advances and adjusted conclusions continually occurring in Gazzaniga's field), then whatever the magnitude of the long reach of genetic inheritances and all of the items outside of one's grasp, each of us is ultimately a creature with agency who does have at least some degree of control over what we do, albeit a measure that may vary widely from person to person and from time to time within one person's life.

Ravven replies directly to Gazzaniga's position on this by citing his lack of consideration of neuroplasticity – that the brains' neocortical pathways are rewired by experience –, and that because of that neural characteristic the influences of culture, meaning, and language are not voluntary but are flexible: we cannot change our patterns of thought through will but 'only by training and re-training'.⁶³ This argument, although it stresses the degree to which the brain is molded by the default forces of culture, meaning, and language into which one is born (without giving, or even being asked for, one's consent), does yet still appear to leave the door of choice open a crack, at least as far as situational interpretations go (and in that, surely, attempting a 'training and re-training'), a factor which would clearly play a large part in subsequent actions.

Yet Ravven on the whole appears to be uncomfortable with such a view, preferring instead the more robust near-determinism that marks her discussions of the changing self, the contexts in which one moves, and the primacy of the ego-centeredness that is an ancient evolutionary heritage fueling the affective systems which in turn drive behavior. This is Haidt's social intuitionism, which Ravven also references (as well as Joshua Greene's work),⁶⁴ with little or no hope for a self-reflective feedback loop – a position that I would

with being unaware.

⁶³ Ravven, *op. cit.*, pp. 470-471, note 96.

⁶⁴ Haidt 2001 and 2012, *op. cit.*, and Greene, *op. cit.* Both of these researchers present the brain's

be very surprised to find either psychologist holding. In her book's concluding section, 'A Final Word on Moral Responsibility', Ravven even goes so far as to state that, 'If free will is relinquished, we come to recognize that what must be *must indeed be*, and that what must have been *could not have been otherwise*.'⁶⁵ This is a very comforting thought to all of us who have regrets, but if true then it would take us into a realm of unalterable fate that not a few would find less than desirable, and it would also disrupt much in regards to the meaning and purpose in life that many people believe they have discovered for themselves.

There is a final problem with this account as well, and that is that Ravven appears to take what are still controversial findings as conclusive. For example, she cites António Damásio's view that the self defines itself by its relations to the environment and its relationships with others (very similarly to how Ravven structures her account, and I should note that Damásio also writes from a biological-scientific perspective)⁶⁶ as being a fundamental and crucial point, yet she also states that the data is at present inconclusive. Panksepp's work is also disputed, as is the extent to which Ravven embraces neuroplasticity. None of this signals that this research should be rejected of course, but as remarked shortly above, when drawing conclusions from the studies being done on the brain it must be remembered that the field of neuroscience, along with its revelations about cognitive functioning, is a rapidly changing one; this is something that Ravven does not always seem to bear in mind. Although her ideas are intriguing, it is my impression that Ravven's contextualized soft realist self would be a far stronger and more applicable self theory if it were tempered down by a large margin, if she sought only to emphasize the

functioning as being prewired and specialized by area – the modular view that is at present generally accepted but which Ravven rejects in favor of a more plastic view.

⁶⁵ Ravven, *op. cit.*, p. 419; emphases in the original.

⁶⁶ Damásio's referenced works are *Descartes' Error: Emotion, Reason, and the Human Brain* (New York: Putnam, 1994) and *The Feeling of What Happens: Body and Emotion in the Making of Consciousness* (New York: Harcourt Brace, 1999). Damásio's more recent work on the self will be explored in our own study below.

enworlded⁶⁷ reality of the self and did not thereafter take the further deterministic steps that she does.

1.2 Answers? – Guiding questions

In the next chapter we will take much from what we learned through our survey of these accounts, particularly from Kristjánsson's soft realist self, but also from Ravven's thoughts as well, in our offering of a self concept that seeks to accurately reflect the conclusions about natural human functioning that are widely accepted, and that may allow us to approach ourselves, and one another, in a more affirmative, enlightening, engaged and engaging manner. In so doing we will try to answer, at least initially, some of the more pressing questions regarding the self: What can the self be said to be and what is it comprised of? What is the relationship between the psychological sense of self and the physical body? How is the self situated in its environment and amongst other selves? Only by first arriving at a self-view that can incorporate both our everyday lived experiences of person and world, and of what modern empirical research can tell us about how the brain is thought to work and where behavior is thought to originate, will we be able to have a clear enough picture to extend our findings into areas of application. If we wish to live and relate better – with those around us, yes, but not least with ourselves – then we will need to begin at that point from which we engage all that we meet within and without: the self.

⁶⁷ Deeply embedded, including perspectival (i.e. perceptual and conceptual) areas. I am grateful to Professor Saulius Geniusas of the Chinese University of Hong Kong for introducing me to this term.

Chapter 2: Laying the Groundwork: Psychological and embedded factors, proposing an alternate soft realist self theory

2.1 Situating

After our consideration of some of the most pertinent background issues and representative self-theories in the previous chapter, I believe that a case can be made that there does seem to be good reason for accepting the commonsense view of a substantive (in some way) self, although that substance is almost certainly not in any kind of physical or definitive form. Even the soft anti-realist account, after all, is based on what amounts to a realist assumption – with its something supported by nothing – despite its expressed non-realist outlook. (The hard anti-realist position is of course the major ‘absolutely not’ non-substantive exception here). In this chapter then we will endeavor to explore ways in which the *how* of all this might play out.

Of the soft realist options we encountered, Ravven’s view of an emotional self made up only or primarily by a biological drive to survival, and one that furthermore exists as a shifting multiplicity dependent on current relational statuses, seemed excessively shallow, and it may also be noted that its foundational survival drive could be included in Kristjánsson’s fuller three sets of self-related emotions (likely in the third category of self-conscious emotions). It was Kristjánsson’s account generally, in fact, that struck us as having the theoretical grounds with the most potential, and I think that the deficiencies found in it can be made up for by supplementing the position rather than by starting over again from scratch. In the below we will therefore attempt to work with this account in order to build up our alternative self concept from out of it, beginning with a wider and deeper exploration of some of the psychological and cognitive issues already raised, then

adding to them, and finally analyzing such in light of a central argument on what a substantive, emotionally-based realist self might compositionally consist of. This process will also include a return to Hume and a more rigorous analysis of human emotions and intuition, aiming to achieve a clearer understanding of the ways these two root biological functions differ and interact. Following that we will expand and reinforce the proffered self-view through considering oft-raised bodily and embodied issues, contextual/embedded considerations, and the roles that uncertainty and randomness play both in our comprehension of the world generally and our place within it. With those issues covered we will hopefully be in a solid enough position to discuss in the following chapter some potential general and phenomenological objections to our self-account, and what some responses and replies to them may be.

2.2 Psychological issues and the self

A. Cognitive structure

This section will start with a review and analysis of the psychological problems at stake in Kristjánsson's soft realist self concept before moving on to the philosophical analysis of such views. I have two reasons for this methodological approach: 1) Psychological research, at least of the kind related here (i.e. non-therapeutic), aims by rigorous and verifiable data-gathering to determine the manner in which we naturally (biologically) function in the world we find ourselves living in, and 2) Ignoring this evidence because we think it irrelevant or because we disagree with the *interpretation* it has been given essentially amounts to committing ourselves to speculation, to the squandering of a valuable resource near to hand. One does not argue philosophically with the laws of physics, one takes them and decides – philosophically – what to do with them. Such 'laws' may ultimately prove to be tentative, they may be misconstrued or misconceived, they may be illusory in one way or another, but they are based on what has generally proven to be an

overall reliable set of procedures for learning about the universe (and especially about what works in the medium scale range of objects we inhabit). I can see no reason not to be empirically informed, though I can see plenty of reasons not to merely take empirically-based conclusions at their word, as it were. While psychology is of course not physics, and while the results of psychological research are admittedly more open to error – or to looser/wider hermeneutic possibilities – than those of the so-called hard sciences, they are at least playing in the same stadium, while philosophy, on the other hand, is across the street warming up for its match against literature. This is not to downplay philosophy’s role, nor to cast aspersions on what it has to offer. Philosophy is, or can be, what holds the sciences and humanities together, giving meaning where there may be only data and teaching rigorous thought and analysis where there may be too much airy dreaming – or too little. As philosophers we can have the best of both worlds, but only if we allow ourselves to be aware of and up-to-date on them.

To begin then, to Kristjánsson’s affective and self-reflective account we may add the growing psychological evidence in favor of an intuition first, reasoning second model of decision-making and behavior initiation, and moreover augment this with Ravven’s emphasis on background causal factors and social pressures playing an important role in shaping the parameters that influence one’s intuitions. It may be remembered that Kristjánsson argues for a one-tiered model of decision-making that grows out of an emotional disposition which a person chooses to adopt due to moral concerns and then, through time, internalizes: his ‘rationally grounded emotion’.⁶⁸ In Chapter 1 we discussed how these conclusions appear to mistake emotions for intuitions, and how the focus on the way(s) intuitions can be affected rather than on how they naturally operate is misplaced. There we found fault with Kristjánsson’s description because it belies what the empirical

⁶⁸ Kristjánsson, *op. cit.*

evidence indicates in favor of a model that, possibly following his Aristotelean convictions, allows rationality to take a central place. While it is true that the moral concerns we adopt are important, they are not the whole of the story, and I believe it is worth dwelling on this point for a short time in order to try and untangle how intuitive judgments and rationality interact, how they come together, and at what points they split apart.

If we accept Jonathan Haidt's model of (potentially) socially-shaped intuitions driving much of our behavior through automatic, affectively-based reactions,⁶⁹ we will think that Kristjánsson has simply got it wrong, as mentioned. This judgment, however, might be premature if we recall that Kristjánsson's aim in the training of a moral self as discussed in the previous chapter was grounded in a concern for personal ethical merit, and so while it is admittedly not entirely clear what Kristjánsson is proposing, I think it worth taking a closer look at his ideas if for no other reason than their relatedness to a soft realist self position. When we do so we can see at least two possibilities emerge.

For one, Kristjánsson does indeed seem to mean the building of one's emotional dispositions out of a root rationality, in a similar way to how one might attach prostheses to enhance natural strength (or, better yet, as one might add emotion-mimicking programming to a robot in an attempt to make it function better around humans). His theoretical picture *is* therefore partially at odds with the empirical results achieved thus far concerning how we as biological creatures naturally function.⁷⁰ Yet if, on the other hand, Kristjánsson also means, or perhaps means primarily, the choice to foster certain intuitions, to reflect on the type of intuitive reactions one would like to have, and then to actively seek to establish and

⁶⁹ Haidt 2001, 2012, *op. cit.*; Greene *op. cit.* and Gazzaniga *op. cit.* also support the general picture Haidt presents even if somewhat differing in the details.

⁷⁰ *ibid.* All four sources referred to support a model of rationality as a later and less important (evolutionarily speaking) development than the automatic emotional responses.

invigorate those intuitions, then his model would in that sense more or less match Haidt's, and it would additionally fit with the results of the scientific research that are involved. Haidt's Social Intuitionist Approach does contain after all, and as an important element, the external feedback loop into one's intuitions (best understood as automatically-generated judgments regarding perceived/experienced stimuli), and this element is composed of both self-reflection and social criticisms and influences. These factors are thought able to alter, build, create, and/or eliminate intuitive reactions during the course of one's life, and they are necessarily based on the experiences and interactions of the person involved. Put another way, on the structure Haidt presents, intuitions can and often do change, and such internal transformations are accomplished not only through the efforts of others but also through one's own (although the 'efforts' of others are more commonly unintentional, coming in the form of unplanned influences rather than outright attempts to remake long-term behavioral patterns). Might Kristjánsson be both right and wrong?

Note, though, that if Kristjánsson did intend to emphasize the latter – thinking about and trying to alter existing 'emotional dispositions' (by which he appears to mean preferred automatic reactions; as discussed such are more accurately labeled intuitions) – then his model is no longer of the one-tiered form that he insists on; it is rather two-tiered: 1) intuitive reaction yielding behavioral output (the expression of the underlying 'emotional disposition'/intuitive judgment), and 2) behavioral output deemed unsatisfactory after its expression and recognition by the individual and henceforth a rational decision is made to attempt to amend future behavior by changing the 'emotional'/intuitive foundation. It will be apparent to the reader that this is the same two-tiered structure that the psychological research has given us. Kristjánsson's self, as the emotionally-based day-to-day psychological unit that he describes it as being, is something that can be formed and shifted during a lifetime, just like the intuitions which trigger behavioral responses can be formed

and shifted through efforts at monitoring personal behavior with one's purposely determined goals for optimal actions kept in mind. According to the research done by Haidt and others, we can work at giving ourselves intuitive reactions out of which different (preferred) results can spring, but the model being represented by such science is still a two-tiered one. Kristjánsson, if understood in this way, has not mistaken the psychological landscape so much as he has just drawn an inaccurate map of it.

Other recent work done in psychological and cognitive studies also supports a view of the self that more closely aligns with a soft realist concept such as that explored above. In a priming experiment gauging self-face reaction times, Steven M. Platek *et al.* found that showing subjects their name printed out, heard via a recording, or when their own body odor was smelled, gave them 'enhanced self-face processing and promoted significantly faster self-face reaction times, suggesting the existence of an underlying, unitary self-processing system.'⁷¹ The authors add that 'the fact that presenting information about the self in other modalities [of the brain] affects self-face identification clearly shows that self-recognition is an expression of an integrated, multi-modal self-processing system.'⁷² Lest these terms mislead us, it is important to note here that the 'self' being referred to is not the soft realist self we have heretofore been considering, rather the 'self' of the self-face reaction times study is the 'self' that we see in the mirror, the whole person self whose core (if such there be) we have been trying to discover. The comparison is not at all apt, but if we are thinking in terms of the self as the pilot of the ship, then our soft realist self would be the pilot and the self-face study 'self' would be the ship seen with the pilot at the wheel. (I would like to caution here that by no means do I intend to advocate the pilot

⁷¹ Gordon G. Gallup, Jr, James R. Anderson, and Steven M. Platek, 'Self-Recognition' in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), pp. 80-110 (p. 88); the article references this study: Steven M. Platek, Julian Paul Keenan, Gordon G. Gallup, Jr, and Feroze B. Mohamed, 'Where am I? The Neurological Correlates of Self and Other', *Cognitive Brain Research*, 19 (2004), 114-122.

⁷² *ibid.*, p. 88.

and ship dualist view by making use of this analogy, it may simply be easiest to see the difference in referents using an example along these lines.)

Nevertheless, what this research suggests is that the manner in which we understand what we are is likely based on function specific areas that are spread throughout the brain (fitting with the modal model presented by Gazzaniga and others), a result also supported by Kai Vogeley and Shaun Gallagher's work, who write that the self, like consciousness, seems to have 'no candidate region or system' that could be considered as generative, no single self-specialist region or module, in other words.⁷³ The objection may legitimately be raised that the above studies shed little light on much other than on how individuals can be prompted to better (more quickly) recognize their own whole persons, but it should be remembered that what we are considering at this point in our study is the psychological side, and certainly from that perspective (psychologically) all of us do have a notion of what we are in terms of self, whether that is able to be clearly verbalized or not. A faster recognition of that self, therefore, indicates a more efficient data processing time by the brain, which in turn can (or at least may) add potentially useful information to the notion of the self as a 'day-to-day psychological unit' available to awareness upon reflection and subsisting at some mental level when not under deliberation. As far as issues pertaining to composition are concerned, I would agree that research like the immediately related does not offer a great deal, but it does yield more, it is another piece to the puzzle, something else to bear with us as we carry on. At any rate, outright dismissal of such data would surely be overhasty.

What these results hint at, moreover, is that one's internal sense of self, a sense that is a

⁷³ Kai Vogeley and Shaun Gallagher, 'Self in the Brain' in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), pp. 111-136 (p. 114).

result of an interacting cognitive system that operates in a unitary manner but which is composed of multiple mental modules working together, may be an emergent property of the normal functioning of the brain, perhaps in a way similar to how the mind itself is thought to be an emergent property of the multiple consciousness systems.⁷⁴ This idea – and again at this starting point we stress that it is really more of a conjecture – helps us to begin to understand how the self might be real in some form of a soft realist way, that is, how it could actually be a ‘something’ yet without our being able to further specify exactly where it is located or what its precise physical/chemical make-up may (or may not) consist of. It anyway gives us something to go on: a toehold, as it were. It must be stressed that this research does not and cannot tell us what the self specifically is, and the fact that it is all too easy to conflate words like ‘self-recognition’ and ‘self-processing’ with ‘self concept’ or ‘self-view’ goes a long way towards explaining these traps and to how the flourishing of self research in recent decades has generated its own share of confusion. It sounds too a cautionary tone that we would do well to heed in our current examination. Care must be taken at all levels and points.

We note, finally, as a last small word of support for the above and as a precursor to Section 2.3 below on bodily/embodied issues, José Luis Bermúdez’s highlighting of the notion that although we view the world egocentrically, there is no specific body point that is ‘me’, rather it is the whole.⁷⁵ We experience our surroundings as embodied beings, and thus any study of the self would be incomplete without accounting for the bodies that we are or have. (That I can write, at this point, about the body as something that ‘we are’ or ‘we have’ tells us immediately how crucial a clear conception of this aspect of the self will be.)

⁷⁴ On the mind, see Gazzaniga, *op. cit.* Issues of mind and consciousness will also be discussed in much more depth below.

⁷⁵ José Luis Bermúdez, ‘Bodily Awareness and Self-Consciousness’ in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), pp. 157-179.

B. Going back to Kristjánsson's sets

What does this mean for the self? Is the self – separately – anything at all? If so, what? In thinking about this problem in light of the foregoing it will be helpful to first remind ourselves of what the contents are of Kristjánsson's three sets of self-related emotions that comprise his realist self:

1. *Self-constituting emotions*: those that define, one's 'core commitments, traits, aspirations or ideals'
2. *Self-comparative emotions*: those that take the self as 'an indirect object' or 'a reference point' for 'comparison with a baseline of expectations'
3. *Self-conscious emotions*: those that are in the self they are about, that take the self as 'their direct attentional and intentional object'⁷⁶

Here we will initially need to discuss Set 1 on its own as it lays the groundwork for the further two sets and as it is the set most open to change through one's thoughts and decisions. We note first that 'core commitments, traits, aspirations or ideals' are not really emotions at all, and further that they are the type of characteristics that we expect to flow out of genetic influence and early experiences. For example, one is not likely to aspire to a career as a professional basketball player if one is very short and naturally uncoordinated, and if one has been taught from a very young age that one ought to give regularly to charities or that one's life is best spent through the service to others then one will likely have adult commitments that reflect that. In calling such 'emotions' Kristjánsson may have meant that the emotions we have *about* our commitments, traits, aspirations or ideals are essential to one's sense of self, but if so that is a different claim from the one that states

⁷⁶ Kristjánsson, *op. cit.*, pp. 75-77.

they are compositional of the self. Labeling these items as compositional is a deeper assertion, a substantive assertion, and given the set's place within Kristjánsson's scheme it appears that is the level intended. Which, I think, is both fine and right – these features need not technically be emotions to have a place within his structuring and, more importantly really, a place within a robust self-view. Additionally, however the phrasing 'self-constituting emotions' is interpreted, the characteristics indicated are significant and formative of the intuitions we come to have (and thereby out of which behavior is generated and decision-making is affected). I will present my own case for using these sets for a weaker version of a soft realist self in the section below on selves, personal identity, and whole persons; in this section I wish only to illustrate how such might be used in a more strictly Kristjánssonean way that also incorporates the cognitive structural elements discussed immediately above.

All three sets are important for the type of two-tiered mental model being described, a model where the base level generates most of what one does but the second level is able to modify and develop the first. To begin, Set 1's self-constituting elements can be understood to include a part (a large part?) of elements that one is more or less born with, that are genetically and randomly determined by who one's parents were, sociohistorical and economic settings, formative experiences during early childhood, et cetera.⁷⁷ Despite, or even because of, the way one's past reaches out to the present, items such as commitments, aspirations, and ideals can of course be changed either through what one exposes oneself to (via, say, learning) or through what others successfully argue one into adopting (Haidt's external feedback loop; an internal reflective feedback loop is also proposed in Haidt's model). Sets 2 and 3 (self-comparative and self-conscious emotions) then allow one to keep track of how one is doing vis-à-vis the type of person that one thinks one wants to

⁷⁷ This point will be more fully formed and discussed in the below.

become. However unclear the extent to which such self-transformation is possible (Gazzaniga does describe the brain as being mostly automatic and unreachable⁷⁸), it is nevertheless there, as lived experiences can testify to. If we wish to shape who we are – rather than simply being shaped by where and when we are – we can, and we do so through the concerted and sustained efforts required to give ourselves new affective bases, new intuitive judgments and automatic reactions by observing our own behavior and working towards thinking (rational reflection) and acting (considered, second mental tier (on our two-leveled model) choices) in ways we think preferable. We not only *decide on* new ‘commitments, traits, aspirations or ideals’ that we would like to have but we actually *take them on*, we incorporate them into our core (emotional) selves. Such is accomplished through repetitive practice of the ‘self-improvement’ variety, e.g. when a co-worker has done something deemed offensive (intuitive reaction), one internally tells oneself that she did not mean it that way, that one’s angry or hurt feelings are likely misplaced, that it is all probably a misunderstanding, followed by an action such as going to the co-worker to discuss the incident in question and (calmly) exchange words on how it was perceived and received. If a process such as this is gone through enough times it amounts to a re-training of the intuition(s) involved, a rationally induced check on the pre-rational anger/hurt that tapers and blunts to the degree that, done enough, results in an alteration of the initial reaction. How much is ‘enough’ here? I think that is very hard to say, and I can offer no more of a response than to state that it very probably depends on the individual and whatever Set 1 elements the person has. Kristjánsson has described this same procedure on his single-level mental model as a part of the construction of a moral self, and indeed it is, but for reasons already outlined we have found the dual-level mental picture to be more accurate and to fit with this type of exercise better, or at least just as well.

⁷⁸ Gazzaniga, *op. cit.*

This then, these three sets in their whole, is the psychological structure of the self that Kristjánsson has proposed as being realist in the soft manner described. It is an account that posits that the self is composed of the emotions one has about who one is, and that the manner in which these emotions work together forms a ‘psychological unit’ that can be said to be ‘the self’ (think of Hume’s ‘bundles’ here; what Hume actually said will be examined in the next section). As with many things written about the self, we find that we have to stretch our imaginations to see what this could really entail, and that in some ways the position is still unclear as to what the internal perspective and voice with which we are all so familiar could be and could mean. It is particularly unclear how this might differ from personal identity, but that Kristjánsson thinks it does differ seems likely in the emphasis given to the changing emotions as the fundamentals out of which the self is born: both Sets 2 and 3 serve dual referencing and compositional roles, whereas for Kristjánsson personal identity’s role appears to be purely as an internal referent (the type of person I think I am, or who/what I see myself as being). However, this judgment is only one that is inferred as Kristjánsson is not clear on this point; I admit that and want to stress it for my inference may well be mistaken.

Going forward, however, it is paramount to be clear about these terms when working with this self concept because while ‘personal identity’ can often be used interchangeably with ‘the self’ (especially in non-realist accounts) it does not have the compositional element required to be able to equate Kristjánsson’s realist self with personal identity – personal identity describes but does not construct. That is to say, if the self is real and separate, made from the three emotive sets that Kristjánsson proposes, then personal identity must actually be generated by the sum of all of those three sets (meaning too that in the case where *only* the three sets were sufficient, important elements such as bodily issues and others would be left out), but again on Kristjánsson’s self-view we instead have the sum of

the sets as being *the self*, and mention of personal identity is missing. This raises the question of whether or not Kristjánsson actually equates the self with personal identity (Set 1 + Set 2 + Set 3 = the self = personal identity). In self concepts where the self can be thought of as personal identity, such as Lockean or neo-Lockean accounts like Parfit's, personal identity is guaranteed by psychological continuity, and so such an equation as that just given would not be an issue. Yet for Kristjánsson's self concept (if he does see it this way), that would not work since one of the points in making a claim for a realist self, even a soft realist one, is to argue that the self is an entity or function (depending on how realist one wishes to be) that is more foundational and that ensures personal identity continues even when psychological continuity does not. Kristjánsson, at least, seems to want more than the Lockean perspective would allow – and indeed his account demands such – and thereby to be placing the self at a deeper psychological level than personal identity inhabits. If so then this view of the self as being composed of the three sets listed could be improved or filled out by having their sum feed into (but not be equal to) personal identity; this is a point we will revisit shortly. Returning now to Kristjánsson's sets purely as he has them, we can note that Set 1 seems to carry the heaviest identity burden and that Sets 2 and 3 are emotions regarding how one feels about oneself (including what one does) referencing Set 1, and also how one feels about what happens to one in the world (including how others react to one) in light of Set 1.

It may be objected that all this is overly vague and too unsteady an account to provide the basis for a realist self, particularly when we consider the extent to which our emotions are constantly in flux. Moreover, what seems to be carrying the real explanatory burden here is Set 1 – and that only. Such a criticism however overlooks that by Kristjánsson's self concept it is the interaction between Set 1's foundation and Sets 2 and 3's interplay with that foundation that together give rise to the sense of self. Its indefiniteness is thereby

arguably a strength in the flexibility it allows and amorphism it tolerates, yielding a self as in flux as its emotional contents. Nevertheless, in this same indefiniteness and lacking a more detailed explanation of the label affixed to the self of a ‘day-to-day psychological unit’ Kristjánsson’s self-view does fail to make satisfactorily clear – at least to me – what is meant by ‘self’ in the (soft) realist sense he claims, and nor does it differentiate the self from personal identity in the full way that his insistence on the affective compositional nature of the self appears to call for. In the course of our study we will therefore continue to build on this account and to branch out from it as we attempt to address these and related concerns. At this point the reader may also be wondering where the continuation of the self could be if it is not to be found in psychological continuity; after all, again as a realist position Kristjánsson’s needs to give an accounting for such, even if his is a ‘soft’ – and therefore nonphysical, non-pinpointable – realism. I will argue that this view of the self, or anyway our altered take on it, does have psychological continuity (but understood differently), and moreover that it finds its basis for continuity and realism in the emotional investment, uniquely held identity, and physical constancy that contribute to the day-to-day psychological unit described (and I will try to fill out what that ‘unit’ might be and/or mean). On the self-view to be given, one does not have a series of separate selves nor does one create new ‘psychological units’ throughout one’s life and via the efforts one exerts, rather one makes adjustments to varying degrees at varying times to one’s already existing (senses of) self. We maintain the same day-to-day psychological unit, the same self, all the years of our self-aware lives; what differs does not do so in kind, only detail. This, I will put forward, is the essence of how a soft realist self could be thought to remain in the midst of its many changes.

C. Going back to Hume

I largely disagree with Kristjánsson’s interpretation of Hume, as carefully as he words it

and as open as the original text is; the way I understand Hume's presentation of the self is as at core equating it with personal identity. If so, this stance would not allow the self to have the realist foundation that Kristjánsson gives it since he stakes his assertions on that point as being based on Hume, as remarked earlier. In what follows, therefore, I will try to make my case for this reading of Hume by going back to the end of Book One in his *A Treatise of Human Nature*,⁷⁹ in which personal identity is addressed directly (in 1:4:6), and then looking at all the places in Book Two where Hume refers to the self to try and understand what he may have meant by his use of the term. Prior to examining the quoted sections from Hume's work though we will need to give a brief background on his notion of the 'double relation of ideas and impressions' to try and see where Kristjánsson likely took the basis for his emotional self as it was described in the foregoing.

Hume's system in Book Two depends on this 'double relation of ideas and impressions', in which an idea of something (e.g. Hume gives the example of one's own house) produces an impression (such as pleasure, assuming it is a nice house), but also additionally correlates to another idea (that of my self, the owner of the house), which is itself connected with another impression (the pleasurable feeling of pride in owning something nice); this latter impression, it will be noted, resembles the earlier impression: pleasurable pride and pleasure. The association of concepts here is therefore said to grant to one the idea of one's self, as Hume writes: 'That is, the subject brings to mind the idea of precisely that individual who is always the object of pride.'⁸⁰ It strikes me as probable that Kristjánsson took his notion of an emotionally generated self from the interplay involved in this relationship, and I think so because he specifically references Book Two of *A Treatise of Human Nature* when he writes in his own account that Hume, 'seems to be

⁷⁹ David Hume, *A Treatise of Human Nature*, intro. and ed. by David Fate Norton and Mary J. Norton (Oxford: Oxford University Press, 1739/2000).

⁸⁰ *ibid.*, Editor's Introduction p. 151.

arguing that, whereas the self as a succession of related ideas and impressions cannot be a direct object for the understanding, the self of whose moral actions each of us is intimately conscious can be a direct object for our emotions'.⁸¹ The emphasis Kristjánsson gives is to the emotional connections in chains of relations like the one just related: an idea to a feeling to another idea that connects the two. Kristjánsson seems to say that this is not a process that is rationally comprehended so much as it is emotionally incorporated, that the self underlying the interplay is subtly sensed rather than (internally) explicitly described. We must tread somewhat lightly when drawing any conclusions here though as what Hume is presenting in his 'double relation of ideas and impressions' is not of course a proper theory of self, rather he seems only to intend to present what he takes to be the common element found between a feeling and the various sources of that feeling, linked also to certain notions (of which the self is one) via association. I believe that what Hume understood the self definitionally to be is better described as a non-self or, following Kristjánsson's classifications as we have been, as a soft anti-realist self. My basis for thinking so will henceforth be laid out.

For a clearer picture of how Hume might have considered the self we turn now to what he actually wrote specifically in that regard. In Book One Part Four Section Six (1:4:6 'Of personal identity'), he displays a skeptical general view in the following: 'For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other...When my perceptions are remov'd for any time, as by sound sleep; so long am I insensible of *myself*, and may truly be said not to exist.'⁸² To this he adds, 'they [i.e. 'we', that is, people] are nothing but a bundle or collection of different perceptions,

⁸¹ Kristjánsson, *op. cit.*, pp. 47-48; note here again and throughout Kristjánsson's usage of 'conscious' as synonymous with 'aware'; this topic will be touched on in this chapter and additionally discussed in following chapters, but I believe equating the two terms tends to lead to more error and imprecision than to anything else.

⁸² Hume, *op. cit.*, Text p. 165; emphases in the original.

which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement.’⁸³ Here we have Hume’s famous ‘bundles’ concept, and with his foundation thus set he summarizes his account:

identity is nothing really belonging to these different perceptions, and uniting them together; but is merely a quality, which we attribute to them, because of the union of their ideas in the imagination, when we reflect upon them...’Tis, therefore, on some of those three relations of resemblance, contiguity and causation, that identity depends; and as the very essence of these relations consists in their producing an easy transition of ideas; it follows, that our notions of personal identity, proceed entirely from the smooth and uninterrupted progress of the thought along a train of connected ideas, according to the principles above-explain’d.⁸⁴

What Hume might be giving us here could perhaps on Kristjánsson’s terms be called a soft anti-realist account that ostensibly equates the self (or rather sense of self) with personal identity and that is dependent on psychological continuity; note that without perceptions (i.e. psychological stimuli) one ‘may truly be said not to exist’. Identity is here only an applied label, a name, a ‘quality’ attached due to the ‘union of ideas’ present in our imaginations when we turn our attention there. This label of one’s given personal identity then, by its defining nature, makes us think that there is a self, that we have a self, for whenever we do think of such we always arrive at this label. This supposition is further supported by memory, which discovers and helps produce an identity in the thinking and observing person, is aided by the emotions working with one’s imagination in linking present concerns with past and future,⁸⁵ and finally by the idea of causation which allows

⁸³ *ibid.*, Text p. 165.

⁸⁴ *ibid.*, Text pp. 169-170.

⁸⁵ *ibid.*, Text p. 170.

one to take identity beyond memory and suppose the same identity (since it is causal of the memory, et cetera) to be in existence even during those stretches of time where memory fails.⁸⁶ Having established sufficient doubt in any realist concept of the self (he supposes), and having described what personal identity really is and why it erroneously makes one think there is a self resting beyond it (based on the associative process just outlined), Hume concludes his thoughts with this gem of a proclamation: ‘all the nice and subtile questions concerning personal identity can never possibly be decided, and are to be regarded rather as grammatical than as philosophical difficulties.’⁸⁷

Doubts may however understandably linger about just what Hume actually meant, and so for the moment I ask only that we keep the preceding in mind as we turn to Book Two to see how Hume wrote of the self there and then attempt to ascertain if at heart his indeed could be thought of as an anti-realist account. This is worth examining since if my reading of Hume is right and he does give to personal identity the weight that Kristjánsson wants a realist self to carry, then the self-view on Hume’s account – on its own based solely on what Hume writes about it – may not be able to be ‘a direct object for our emotions’ as Kristjánsson has it, and if that is the case then Kristjánsson might not have enough to cement his self theory in the manner he tries to. On the other hand, if we think that *personal identity* can be a direct object for our emotions, and I do not see why it could not, Kristjánsson’s core idea regarding the three emotional sets and their substance-providing role would still stand, but for the reasons given in the previous section such would nevertheless be inadequate for a realist view of the self. That is, although personal identity might be an object for one’s emotions, it would not thereby generate a self able to underpin personal identity, and on Kristjánsson’s self-view the self, as (soft) realist, would need to

⁸⁶ *ibid.*, Text pp. 170-171.

⁸⁷ *ibid.*, Text p. 171.

be separate from and supportive of personal identity, as argued. Again though caution is in order when thinking about where all this could be leading, for such a conclusion only on its own would not mean that the self is necessarily non-existent (i.e. anti-realist), nor would it mean that Kristjánsson's self concept is without value, it would only mean that Kristjánsson may be mistaken in what he attributes to Hume. If reading Hume did give Kristjánsson his ideas about how a (soft) realist self could be established then certainly Hume deserves the credit given him – I do not mean to object to that –, but perhaps not in the way that Kristjánsson is evidently attributing it. The point might only be academic, but let us nevertheless dig further into Hume's work in an attempt to understand his stance a bit more thoroughly.

The first mention of the self in Book Two comes in Part One Section Two (2:1:2 'Of pride and humility; their objects and causes') where Hume gives us a definition of the self that seemingly just substitutes the term 'self' for 'personal identity' as he previously summarized it: 'Tis evident, that pride and humility, tho' directly contrary, have yet the same OBJECT. This object is self, or that succession of related ideas and impressions, of which we have an intimate memory and consciousness. Here the view always fixes when we are actuated by either of these passions.'⁸⁸ This is succeeded by a statement that shifts the focus somewhat, moving back to perceptions (and not 'ideas and impressions') that closely matches Hume's 'bundles' line quoted above: 'But tho' that connected succession of perceptions, which we call *self*, be always the object of these two passions, 'tis impossible it can be their CAUSE, or be sufficient alone to excite them.'⁸⁹ Following on along these lines of distinguishing between object and cause we find this thought, which appears to be the root of Kristjánsson's reading of Hume:

⁸⁸ *ibid.*, Text p. 182.

⁸⁹ *ibid.*, Text p. 182; emphasis in the original.

Pride and humility, being once rais'd, immediately turn our attention to ourself, and regard that as their ultimate and final object; but there is something farther requisite in order to raise them: Something, which is peculiar to one of the passions, and produces not both in the very same degree. The first idea, that is presented to the mind, is that of the cause or productive principle. This excites the passion, connected with it; and that passion, when excited, turns our view to another idea, which is that of self. Here then is a passion plac'd betwixt two ideas, of which the one produces it, and the other is produc'd by it. The first idea, therefore, represents the *cause*, the second the *object* of the passion.⁹⁰

This seems clear enough: Hume here has the self as the object of the passion (the emotion) of pride or humility which was caused by an idea ('Something', say, the notion of one's house) and then via the related ideas raised by that rests on the idea of the self. Yet the self being described is just that, an idea, incapable on its own of being a cause (and hence the object, bound to the causal 'Something' via the 'passion'). Given this with what Hume previously wrote and was quoted above, to ascertain a realist conception of the self from within this presentation may not be warranted – the self seems too dependent on other items that must *precede* it for it to have the meat and fat that realism requires. Still, I may well be wrong or misinformed in my reading and so we will continue our exploration.

Part One Section Five (2:1:5 'Of the influence of these relations on pride and humility') tells us that 'tis absolutely impossible, from the primary constitution of the mind, that these passions [pride and humility] shou'd ever look beyond self, or that individual person,

⁹⁰ *ibid.*, Text p. 182; emphases in the original.

of whose actions and sentiments each of us is intimately conscious.’⁹¹ The self here suddenly appears to be something more akin to a whole person, but it is possible that Hume still only has in mind personal identity by such a statement since ‘identity’ (referenced by Hume earlier, see the first long quote in this section) was described as the label one gives to oneself due to the conjunction of cause-feelings-idea in regard to ‘me’. It is also possible that Hume did not distinguish between personal identity and whole person, and thus we are reading in theoretical layers that are simply not there. We may note, though, that this is the very statement referenced by Kristjánsson when he gives the basis for his self concept as Book Two of *A Treatise of Human Nature*.⁹² A little later in the same section Hume also describes the self as an idea attached to pride, a ‘contrivance of nature’ like lust or hunger and their objects.⁹³ Yet this is not terribly helpful to us in trying to untangle his view as the objects of lust and hunger are certainly very real (and physically so, in contrast to how Hume has earlier described the self), yet even so the thrust of the argument seems to again be on the notion of an idea necessarily attached to a passion, a resultant object, and that we cannot have the one without the other.

The picture becomes murkier as we find Hume, in Part One Sections Six and Nine, making two further rather cryptic remarks about the self: in 2:1:6 (‘Limitations of this system’), he contrasts pride with joy, the former having two objects necessary to its production (the cause of pride and the self) and the latter having only one (the cause), although he admits that joy must have ‘some relation to self’ – but such is only needed to ‘render it agreeable; nor is self, properly speaking, the object of this passion.’⁹⁴ These statements appear to hint that joy can exist in the absence of a self, which might appear confusing but if Hume’s

⁹¹ *ibid.*, Text p. 187.

⁹² Quoted above: ‘[Hume] seems to be arguing that, whereas the self as a succession of related ideas and impressions cannot be a direct object for the understanding, the self of whose moral actions each of us is intimately conscious can be a direct object for our emotions’, Kristjánsson, *op. cit.*, pp. 47-48.

⁹³ Hume, *op. cit.*, Text p. 188.

⁹⁴ *ibid.*, Text p. 191.

account is non-realist about the self then that outcome would be possible while maintaining that the joy must be experienced by some living creature and not merely by some thing (even though that enjoyer, the object, is not required for said joy's existence by this account; the enjoyer is only necessary for the experience of the existent joy). Then in 2:1:9 ('Of external advantages and disadvantages') Hume writes that, 'pride and humility have the qualities of our mind and body, that is *self*, for their natural and more immediate causes'⁹⁵ (he thereafter notes that there are other causes which obscure the primary), seeming to now introduce the body into the concept of the self, which, if it was important for Hume, would be something else entirely from his earlier 'bundle' of perceptions or series of connected ideas explanations. This, however, is an isolated comment and is not repeated; it is moreover possible to simply read 'identity' where Hume writes 'mind and body, that is *self*', especially if for Hume there is no core self to speak of and personal identity and whole person are not concepts that he separates. However, in Part Two Section One (2:2:1 'Of the objects and causes of love and hatred') Hume does repeat an earlier self definition, from 2:1:5, that, 'the *object* of pride and humility is self or that identical person, of whose thoughts, actions, and sensations we are intimately conscious'.⁹⁶ Here again is our whole person and/or personal identity; just what are we to make of all this?

Taken with Hume's earlier description of what personal identity is (an attribution we assign to our continuing perceptions (see again the first long quote above)) and how perceptions generate the sense of self, which when absent ('as by sound sleep') leads to a condition equated with non-being (see the first quote from 1:4:6 above), all this together gives a picture of the self that is either an illusion generated by perceptions (= personal identity and/or whole person), or the flow of other ideas and impressions relating to the person who

⁹⁵ *ibid.*, Text p. 198; emphasis in the original.

⁹⁶ *ibid.*, Text p. 214; emphasis in the original.

is their object, or is simply the full-package individual ‘of whose thoughts, actions, and sensations we are intimately conscious’ (the ‘we’ here evidently being ‘one’ as in ‘one would like Hume to be more clear’). My own view is that judging just from the space and context Hume gives to each wording, whole person was probably not a notion that he distinguished from personal identity, and in both senses of the self as an illusion (via perceptions or via the train of other ideas) the view of the self he presents is anti-realist and tied to psychological continuity. A realist self, as we have indicated, would require more.

I should stress that this is only my reading of Hume and that it is difficult to know exactly what he meant, but when what he does write about personal identity and the self is laid out in this way it seems clear that Kristjánsson’s interpretation is less able to support what he states he would like it to than what he claims. Granted, even if Hume’s self is an anti-realist one, the sense of self that develops in the way Hume describes could still be an object for emotional activity, and that it could be so does lend credence to Kristjánsson’s three self-related emotional sets, and to the idea of extending those sets. Kristjánsson too may have simply wished to indicate that he used Hume as a jumping off point, but even so the problem of how the self stands in relation to personal identity in Kristjánsson’s self concept remains, especially given our conclusion above that in Hume’s account the terms appear interchangeable (and linked to psychological continuity). The next section will therefore discuss how a soft realist self could underpin personal identity. That is, what it would have to mean for a realist self based on Kristjánsson’s three sets to be ‘real’.

D. Selves, personal identity, and whole persons

In this section I want to say something about psychological continuity that may be somewhat surprising, about how the self relates to personal identity, and about the role that the self plays in relation to the whole person. I will be considering each of these terms

separately. By saying these things I hope to lay the necessary groundwork for the psychological side of the self that will then allow us to examine some (non-psychological physical) bodily issues involved, pausing only along the way in order to get clear on how emotions and intuitions fit into the picture as they lie somewhat between the more focused mental side that we have been discussing and the more purely ‘physical’ side that we will discuss. (I put ‘physical’ in quotation marks/inverted commas because although naturally all mental/psychological phenomena do take place physically Descartes’ legacy has left us with a default dualistic mode of thinking about such; it will later be argued that however we may consider them conceptually these elements do have discoverable physical (biological) foundations.) The structure of our discussion below will have emotions and intuitions examined in the next section, followed thereafter by bodily and embodied issues in the section after that. I think that however integrated we as living organisms may be it is far more helpful for comprehension to take these pieces one at a time.

To think about psychological continuity in a different light than the one Parfit offered (where losing relation R and then taking it up again can be thought of as being like different lives), or the one Hume offered (where deep sleep (the absence of perceptions; that is, psychological stimuli) is like non-being), we will need to consider what it is that makes any kind of psychological continuity possible at all: (conscious) awareness.⁹⁷ Without awareness we could have no ‘psychological connectedness and/or psychological continuity, with the right kind of cause’ (relation R),⁹⁸ nor any perceptions of any kind. The brain would still be active in the way (or ways) it is active for patients in a state of coma – perhaps we might even still dream – but clearly if we do not have awareness then

⁹⁷ I write this as ‘(conscious) awareness’ rather than just as ‘consciousness’ because I will later argue that consciousness remains with us throughout life, regardless of conditions of awareness (in the sense of ‘consciousness’ as it is used in common speech). This topic will be taken up in depth in Chapters 4 and 5.

⁹⁸ Parfit, *op. cit.*, p. 262.

none of our other higher mental functions could be thought to be working, at least not in the ways we are familiar with as the types of creatures we are and on which the above self accounts are based. This all seems straightforward enough. Where I judge that we fall into error, though, is in the way we think about what awareness is, and by this I mean the manner in which consciousness is unreflectively understood, for the two terms (awareness and consciousness) are typically taken to be interchangeable. I do not think that they are – I think rather that consciousness as a condition of the living brain supports and secures both aware and unaware moments –, and although our main analysis of consciousness per se will take place in Chapters 4 and 5, here we will give an outline account of contemporary neuroscientific findings on human consciousness with the aim of understanding what is physically involved when it comes to awareness and to consciousness as we daily experience them.

As previously described, the prevailing model of consciousness among cognitive scientists is that it is a network of specialized consciousness systems that process the contents of their various domains locally but which are linked together in a complex manner, the ‘constellation of consciousness systems’, which may appear unified from the inside (as it were) due to the functioning of an integrative and interpretive cognitive module.⁹⁹ This means that when we speak or think of ourselves as fully conscious when awake (i.e. aware; again, I object to the standard usage of ‘conscious’ but see no way around writing in this manner when addressing the status quo), and fully unconscious when asleep (i.e. unaware), we are mistaking our (theorized) interpretive module (which acts in yielding the sense of being presently aware) for the entire consciousness network. This is an easy distinction to miss since the interpretive module (or function or (possibly) collaborative modular working) provides each of us with that inner voice and perspective that we know so well,

⁹⁹ Gazzaniga, *op. cit.*

and our use of ordinary language phrases such as ‘be conscious of’ reinforces the view that consciousness equals awareness. I would suggest though that instead of this view, based on the modular understanding of consciousness just given, we think of ourselves to *always* have consciousness whether or not we are aware via the interpretive module (or its functionary equivalent). That is to say, if at least one of the very many consciousness systems we have is doing something then we ought to consider our consciousness as ongoing. This apprehension, if right, potentially means that consciousness is continually being sustained all throughout life, and if that is so then psychological continuity is continually being sustained as well. We only think that deep sleep stops or limits or erases this continuity (à la Hume; relation R is more sophisticated in that it adds psychological connectedness (we wake up and are aware of our pre-sleep lives)) because in those states we assume ourselves to have lost all consciousness due to a lack of perceptive cognizance. Once more though, this assumption is based on what I take to be an erroneous conceptual position that equates awareness with consciousness; it is more accurate, I think, to place awareness as a frequent but non-continuous result of how the human (and surely many nonhuman animals’) underlying consciousness systems function.

What is really going on in the brain during such times of disappeared awareness, in the midst of, for example, a state of deep sleep? In addition to its normal maintenance and monitoring activities, the brain during sleep fluctuates between various levels and degrees of consciousness, consolidating memories and producing hormones at different stages along the way.¹⁰⁰ It may be wondered if there is ever any moment during such periods when the entirety of the consciousness network ceases operating all together. Perhaps in

¹⁰⁰ Matthew J. Edlund, ‘Sleep and Your Different States of Consciousness: Thinking may not be what you thought’, *Psychology Today*, 13 January 2011.
<<https://www.psychologytoday.com/blog/the-power-rest/201101/sleep-and-your-different-states-consciousness>>.

the deepest part of a deep sleep? The research seems to point against it, but given the nature of the area of inquiry and our own objectives in this study I am comfortable simply stating that I do not know, and I am also unsure whether anyone could give a fully definitive answer to that. What seems most salient for the present is that the model of many separate interconnected but function-specific areas of consciousness that give rise to an emergent and somehow (felt to be a) unified whole at least makes it possible, and probably plausible, that every single one of them will not stop performing all at once. If that is the case, there is always and necessarily some degree of psychological continuity (continuous consciousness, even if not continuous awareness). While this conclusion in itself is an interesting proposition, it is still not enough to account for the realism of a realist self or even for personal identity. We will consider the former first.

As already remarked, for a realist self to have any meaning it would have to exist at a deep enough psychological level that it could *inform* personal identity and not simply be interchangeable with it (as non-realist accounts sometimes have it). However, in the absence of a hard realist conception of the self the role that the self plays in the greater fullness of the person will necessarily remain somewhat limited when contrasted with the types of claims made on it by views such as those of a traditional soul; the self theory that I wish to put forward is of a weaker version than that (and this is a caution on how ‘realist’ should be read): it is of a self that is just one of the elements that makes up a whole person. Thus, following Kristjánsson’s lead and making use of the elements of his account that we have so far found compelling (the ‘day-to-day psychological unit’, the three ‘emotional’ sets, the focus on the importance of emotions, and the notion of a soft realist self as sufficiently realist), I will attempt to present a realist self concept that is far less than a Cartesian ego but that nevertheless does indicate the importance – the cruciality – of the self insofar as it informs a life.

Something else Hume wrote in Book Two of his *A Treatise of Human Nature* made me immediately think of Kristjánsson's three sets of self-related emotions: in Part Two Section Six (2:2:6 'Of benevolence and anger') Hume compares ideas to solid matter and impressions ('especially reflective ones') to 'colours, tastes, smells and other sensible qualities'.¹⁰¹ As such, ideas cannot be blended, only conjoined, but impressions may be mixed to the extent that they lose their individuality in the new whole; 'Some of the most curious phænomena of the human mind are derive'd from this property of the passions.'¹⁰² Might we therefore see Kristjánsson's Set 1 as an 'idea' and Sets 2 and 3 as 'impressions' in this Humean sense? The shift in thinking such a view would grant could help clear up some of the areas found wanting in Kristjánsson's self concept just as he has it. We noted above that Set 1 is not really an emotional set (except by an unusual interpretation of 'self-constituting emotions'), and that it carries a heavier formative burden than the other two sets, but that the characteristics it highlights become decisive factors contributing to the self. As suggested, then, we might place Set 1 as an unmixable but still alterable 'solid' slightly apart from the other two sets – which are in turn understood as blendable, fluid, and subtle 'senses' – in our consideration of all three. To review, here again are Kristjánsson's three sets:

1. *Self-constituting emotions*: those that define us, our 'core commitments, traits, aspirations or ideals'
2. *Self-comparative emotions*: those that take the self as 'an indirect object' or 'a reference point' for 'comparison with a baseline of expectations'
3. *Self-conscious emotions*: those that are in the self they are about, that take the self as

¹⁰¹ Hume, *op. cit.*, Text p. 236.

¹⁰² *ibid.*, Text p. 236.

‘their direct attentional and intentional object’¹⁰³

On this apprehension Set 1 consists of our preferences and outlooks, the results of our genetic inheritance, our upbringing, and the considered choices we make; these personal characteristics are clearly substantial aspects of our identities but are generated out of limited formative options (e.g. we cannot choose our genetics nor our upbringing). Sets 2 and 3 are our affective and intuitive reactions, responding to the whole of which they are parts and through these responses continually contributing to maintain, shift, sustain, or otherwise adjust this core psychological unit. Kristjánsson has these three sets as composing his soft realist self concept (before our tweak that somewhat separates Set 1 and considers it differently from Sets 2 and 3); the problems of this rather bare account have been examined above. In light of the foregoing, what needs to be changed to improve this self-view and bring it in line with the other details we have examined?

It was argued in our examination of the hard anti-realist position in Chapter 1 that uniqueness can never really be eliminated, and as a result we may wish to add that element to our adjusted Kristjánssonean three sets. It is true that there the uniqueness argument was made in the context of personal identity issues, but it will be remembered that at the time we were considering Parfit’s self concept, which is a hard anti-realist account and, being anti-realist, uses personal identity as synonymous with the self (there being no ‘real self’ to speak of). If uniqueness cannot be removed from personal identity then it stands to reason that it cannot be removed from a realist conception of the self either, because in any robust realism the self needs to be positioned more deeply and foundationally than personal identity, and then from out of that layer in turn helping form one’s personal identity. This much is clear enough, for how could a non-unique self amount to a unique personal

¹⁰³ Kristjánsson, *op. cit.*, pp. 75-77.

identity? Even if, for the sake of argument, it were to be granted that a non-unique self could conceivably form a unique personal identity through some form of variant combinational variety or varieties with certain other items that went into a personal identity, the possible syntheses would only be so many (depending on the number of variables involved and the manner in which they are able to be combined), such that sooner or later a double would result – and hence represent a non-unique personal identity. Admittedly, depending on the variables and combinational possibilities, the chance of such occurring might be small, but it would not be zero. Moreover, on our proffered account the self is necessarily (and not only sufficiently) more fundamental than personal identity, and so it must also have uniqueness as one of its criteria if its (resulting) personal identity does.

The self must too, it seems, have consciousness as one of its primary elements, but this hardly appears to require argument since on the above understanding of consciousness as a continually ongoing process (for at least one of the compositional modules in the consciousness network) a non-conscious self would be a self that has passed away (died). Additionally, always having consciousness means also always having psychological continuity: a never-conscious self would be something else entirely, and the notion strikes me as an impossible claim to make; without ever having achieved even a moment of consciousness none of the self concepts we surveyed would consider a self (or even a perceivable personal identity) to have existed. More interesting to consider is what a non-conscious self would be on the traditional understanding of consciousness, where consciousness is equated with awareness, and there we find *qua* Hume, Parfit, and perhaps Narveson (amongst others) that the self either ceases to exist when not conscious (Hume), or that it requires something in addition (psychological connectedness; Parfit) to make the link back to its earlier conscious time. It is in such cases, however, that we have our anti-realist models, and by them it is not the self that is being discussed but rather personal

identity. In our seeking of realism we have had to draw a line between these concepts of self and personal identity, a line that is not needed for anti-realists as the self is wholly dismissed as a separate entity. At any rate, the traditional view of consciousness appears to be overly simplistic, and so it is perhaps best to remove it from our theoretical thinking.

The final element that we will need for our account of the self will be the body, yet only in a limited sense as here we are focusing on the psychological foundation of the self and setting our discussion of bodily/embodyed issues aside for a later section below.¹⁰⁴ For the time being, therefore, we need only note that our analysis finds a body necessary for a self to exist, and the reader will be asked to be kind enough to table any objections she may have until later. All told, our core psychological soft realist self initially consists of Kristjánsson's Set 1 adapted to add to Kristjánsson's definition of it the sense of core formative characteristics, some of which are indeterminable (e.g. genetics, upbringing, et cetera) but the results of which can still be adjusted (e.g. overcoming childhood trauma); Kristjánsson's Sets 2 and 3, also adapted to add to Kristjánsson's definitions the senses of affective and intuitive reactions/responses to Set 1 and to the self as a whole in an open-ended process throughout a lifetime and grouped together to highlight how they play off and into the more identitarian/labeling and less 'emotional' Set 1; these three sets are then supported and underpinned by consciousness (and therefore also psychological continuity), and the presence of a body (and therefore physical continuity as well; more on that below). Thus: Set 1 + Sets 2 & 3 + C + B_p = the self.¹⁰⁵

How would this view of the self inform personal identity, as it must by the account we

¹⁰⁴ In the following it will be argued that the self arises as an emergent property out of the body's natural functioning in an analogous way to how consciousness (and derivative awareness) arises out of the brain's natural functioning.

¹⁰⁵ 'C' of course stands for consciousness and I use 'B' with a subscript of 'p' to indicate that the emphasis is only on the presence of a physical body (Body (presence)); bodily details will be important for personal identity but as argued that is a step above the self on this view.

have so far presented? Further, in what way would it fit into a whole person? It will be remembered that in this section we are seeking to differentiate all three of these terms. Ludwig Wittgenstein, in his *Blue Book*, writes that how we actually identify and re-identify people depends on a great many contingent facts (physical appearance, habits, voice, behaviors, et cetera), suggesting that personal identity involves much more than a single, important element.¹⁰⁶ This insightful remark can be taken to support the position we have taken that if a self is to be realist in a meaningful but non-Cartesian (or something like a ‘true inner self’) way, then it will be one of the elements that goes into personal identity but is not equal to personal identity (or, to put it another way, is not the whole of personal identity). Although non-realist accounts, as discussed, may not suffer inconsistencies by speaking equally now of self and now of personal identity – as, again, a real self is dismissed –, and although hard realist accounts may even claim that the self is the whole of personal identity – all that a person ‘really’ is –, the soft realist account we are trying to build will have to pass between these two poles and chart a new course. It will need to supply a self that is able to support identity and whole person conceptualizations while being less than both but not so much less that it can be removed. Towards a distinct definition of personal identity then, to the self just described we may therefore add these contingent facts, a part of which importantly entails the bodily elements beyond the mere presence required for the self. These details consist not only of physical appearance and voice but also physical abilities, physical coordination, physical prowess, and the like. The feedback we receive from others will also play into our personal identities, and when this feedback reaches into affecting our selves (via impacting one or more of our adjusted Kristjánssonean three sets) it will be by first going through personal identity as it, being partially composed of contingent facts that exist in the world, is especially responsive

¹⁰⁶ David Bakhurst, ‘Wittgenstein and “I”’, in *Wittgenstein: A Critical Reader*, ed. by Hans-Johann Glock (Oxford: Blackwell Publishers Ltd., 2001), pp. 224-245.

(fragile) to the social whereas the self on our model, being born internally, hinges more on feedback from inward reflection. To put this into an equation, we have: $PI = \text{Self} + CF_B + CF_O + FB_S$.¹⁰⁷

With the self and personal identity defined and distinguished, the whole person is now just a single step away. Our view of personal identity has already given us a grounding and separate/separable psychologically-based self and its social and physical place in the world, all that is further required to fill out the full picture is embeddedness.¹⁰⁸ Thus, to our equation for personal identity we need only add the epochal, historical, geographic, linguistic, socioeconomic, and other elements that combine to situate each person's existence in the wider realm that they inhabit. We hardly require a shorthand for this but $WP = PI + E$ will do just fine.¹⁰⁹

Altogether we now have a self concept that details what is involved in the creation, sustenance, and alterations to a (soft) realist self, how that self plus other factors feed into generating a personal identity, and how both self and personal identity in turn go into the formation of the whole person. Combining the equations (and moving the E variable to the front to make it easier to read) in order to illustrate the hierarchical relationships between these levels we have: $WP = E + PI (\text{Self} (\text{Set 1} + \text{Sets 2 \& 3} + C + B_p) + CF_B + CF_O + FB_S)$. All that remains now in the present section on psychological issues and the self is to get clearer on another important distinction: the one between emotions and intuitions. Once we

¹⁰⁷ 'PI' is personal identity, 'CF' with the subscript 'B' is for contingent facts about the body, whereas 'CF' with the subscript 'O' is for other contingent facts (to emphasize the importance of the body I think it useful to differentiate in this way), and 'FB' with the subscript 'S' is for feedback from the social realm (to distinguish it from internal feedback, which is part of Sets 2 and 3 that go into the self).

¹⁰⁸ It might be thought that embeddedness is not even necessary and that we already have a whole person either from the self and personal identity as described, or just from personal identity in a non-realist sense (if my case has been unconvincing). To see a person this way, though, makes the same mistake of failing to account for context that Ravven highlighted so thoroughly (see Ravven, *op. cit.*), discussed in Section 1.1.D of Chapter 1, 'The contextualized soft realist position'.

¹⁰⁹ 'WP' naturally being whole person and 'E' being embeddedness.

have done that we will be ready to move beyond the psychological and into the bodily/emodied issues that are also significant for the self.

E. Intuitions and emotions

As pointed out above, the distinction between intuitions and emotions is not always clearly represented, nor are the roles they play in relation to our lives – much less the self – as transparent in texts and references as we might like them to be. Intuition in particular is a difficult case as its usage in everyday language is somewhat vague (typically along the lines of one's 'gut feelings'), while its employment in psychological and philosophical literatures at times focuses on different areas, with psychology favoring intuition as a decision-making mechanism and philosophy as either agreeing with that or instead seeing it as a source for (potential) knowledge. To try and untangle intuition we will therefore look briefly at all three facets just remarked on: common usage, automaticity, and epistemology. It will be argued that intuition is best understood as a process or as a composite; such experiences are knowledge-based or knowledge-types of a kind, but they are also preconscious decisions and evaluations stemming from this knowledge.¹¹⁰ Following that we will then consider how to separate emotions from intuitions, and it will be argued further that the former inform but do not determine intuitive reactions to the world we encounter.

What is an intuition? When we speak of intuitions we generally describe such as being hazily noticed senses or impressions, things that are 'just known' or 'just felt'. Although not typically, these perceptions can sometimes be quite powerful, such as when one has a

¹¹⁰ As remarked, for reasons of precision and accuracy I prefer the use of 'preconscious' to the more standard 'subconscious' given the considerations we have seen for an alternate understanding of 'consciousness' continuity, and the differentiation drawn between it and awareness. In some ways the label is still unwieldy for the nuance involved is also akin to 'preaware', but the contents of the preconscious – being automatic – never enter into awareness; what is instead available to awareness, to reflection, are only the observable results of completed preconscious processing.

‘eureka!’ moment and a new and unforeseen insight comes all at once in a flash. Most commonly we hear others say things like they ‘have a feeling about’ this or that; my wife remarked that she had one such feeling that we wouldn’t live in our present (at the time of writing) apartment for very long. As it turns out, she was right: we had moved in just over a year prior to her comment and due to my getting a new position in a city far away we needed to move out again only a few months after that. What was this feeling she had and how is it that she was right? Was it purely a lucky guess or was it something more? Did she have any basis for asserting confidence in her sensed feeling or should she have doubted, ignored, or dismissed it? I do not know if unconditional answers can be given to any of those questions, but that they naturally occur to us demonstrates just how fascinating this aspect of the human experience is.

That intuitions come seemingly out of the blue appears to indicate that they are products of the preconscious (i.e. ‘subconscious’ in traditional usage), and so we might begin our investigation there. One of the more remarkable facets of the preconscious is its processing capacity. While only about seven separate items can be processed in awareness (i.e. ‘consciousness’ as standardly used) at any one time, the entire human system can accommodate 11.2 million units at once.¹¹¹ In an experiment designed to test the veracity of the products of preconscious processing versus those of aware processing or no processing (based on decisions made when evaluated from subjective and normative points of view), Ap Dijksterhuis divided subjects into three groups that determined how they would arrive at their decisions. The preconscious processing group was distracted by an unrelated task for four minutes before being asked to decide, the aware (‘conscious’) processing group was asked to weigh the evidence and think about the options for four

¹¹¹ Ap Dijksterhuis, ‘Think Different: The Merits of Unconscious Thought in Preference Development and Decision Making’, *Journal of Personality and Social Psychology*, 87:5 (2004), 586-598.

minutes prior to deciding, and the no processing group was asked to immediately decide after the options and choice-related information were relayed.¹¹² The preconscious group was found to persistently make better decisions than the other two groups (again, ‘better’ from subjective and normative points of view).¹¹³ In commenting on the results, Dijksterhuis referred to another study which proposed that, in his own words, ‘too much conscious reasoning increases the weight that people attach to reasons that are accessible and easy to verbalize. These reasons are not always the ones that should receive more weight.’¹¹⁴ Dijksterhuis speculates that the preconscious can actively think (though in what manner surely remains unknown; I have not come across even any attempted procedural explanations), that it organizes information more quickly through sorting or grouping, through polarizing (good becomes better, bad worse), or through the continuation of previously assigned/held associations.¹¹⁵ It is informative that this study was based on the evaluations of decisions, for in the above considerations of Haidt’s, Greene’s, and others’ work we saw that intuitions are thought to come in the form of preconscious judgments, and moreover that much conduct is generated from these sources.

Haidt’s Social Intuitionist Model, to review, has it that intuitive judgments (decisions, evaluations) are generated automatically and effortlessly by the situations, perceptions,

¹¹² Some may wish to label what this last group did as making a ‘gut decision’ but such should not be confused with preconscious processing; nor should intuitions, as automatic decisions, be confused with preconscious processing. Where to draw the line on what decisions were intuitive in Dijksterhuis’ study is not entirely clear, however. Is a decision intuitive if it is delivered up following one’s preconscious processing, or only if it comes to one without having been processed at all? If the processing in question occurs preconsciously anyway how would one know whether or not it had occurred once made aware of the resulting decision itself? How ‘automatic’ must an automatic intuition be? Despite these valid concerns what is most relevant for the moment is what this study revealed about the preconscious.

¹¹³ Dijksterhuis, *ibid.*

¹¹⁴ *ibid.*, p. 591; the referenced study is Timothy D. Wilson, Dana S. Dunn, Dolores Kraft, and Douglas J. Lisle, ‘Introspection, Attitude Change, and Attitude-Behavior Consistency: The Disruptive Effects of Explaining Why We Feel the Way We Do’, in *Advances in Experimental Social Psychology*, ed. Leonard Berkowitz (Orlando, FL: Academic Press 1989), 19: 123-205.

¹¹⁵ Dijksterhuis, *op. cit.*

and/or stimuli encountered and that these in turn give rise to behavioral output.¹¹⁶ Many of these reactions will be biologically based and rooted in effective evolutionary reasons, but many others will be the results of cultural conditioning, personal past experiences, social persuasion, and even prior exercises of reasoning that resulted in the creation of new intuitions or the adjustment of already existing intuitions.¹¹⁷ Moreover, regardless of the source of an intuition, the underlying operative structure itself appears to be built in; Osman and Stavy report on their research into the development of children's intuitive systems that intuitive rules once set seem to work like reflexes, to be *prima facie*, and are employed to allow for rapid judgments.¹¹⁸ One example of such an overall process would be that I now, after years of living in Japan, automatically look to my right before crossing the street despite having had it drilled into me as a child in the US to look to my left. I have an alternate intuition running on my extant intuitive substructure, and I do not need to remind myself where to look – or even that I need to look at all – every time I go to cross a street. It may be argued that such is a reflex, or even habit, and not an intuition, but that would be missing the point of what the research seems to be telling us: that intuitions most often work like automatic decisions/evaluations that lead directly to behavior (though let us not forget those 'eureka!' moments). I am approaching a street, there is potential danger, what ought I to do in this situation, et cetera. As mentioned above, in various ways we may also equip ourselves with new or differing intuitions either through purposeful reflection or through the influence of others, whether such is actively noticed or not; we additionally may arrive at new or varied intuitions through affective influences, for the emotions have a strong role to play in what we decide and what we do (to be considered in more depth below). One other notion from the empirical research into intuitions worth mentioning is that of Eugene Sadler-Smith and Erella Shefy's description of intuition as a composite of

¹¹⁶ Haidt 2001, 2012, *op. cit.*; compare Greene's research here as well, in Greene, *op. cit.*

¹¹⁷ Haidt, *ibid.*

¹¹⁸ Osman and Stavy, *op. cit.*

intuition-as-expertise and intuition-as-feeling (or knowing and sensing); they too also emphasize that intuitions can be adapted over time based on novel experiences and the manner in which one deals with such.¹¹⁹

Thus for the psychological comprehension and presentation of intuition, we will now move to an examination of how intuitions have been approached epistemologically with the intention of trying to arrive at a more holistic view. This is necessary because, as the above example of crossing the street hints at, we still seem to be missing something in our description. What may be said of intuition as a way of knowing? What about cognitive acts like my wife's 'just knowing' that we would not stay in our apartment very long? A defense of intuition in this epistemological sense can be found in Elijah Chudnoff's *Intuition*.¹²⁰ There he defends a so-called *sui generis* view of intuition against the so-called doxastic view, putting the situation, as he sees it, this way: 'The simplest doxastic view is that intuitions are judgments. Most doxasticists think that intuitions are a special kind of judgment, perhaps pre-theoretical or spontaneous.'¹²¹ This, as we have seen, is the standard account from the psychological research, and Chudnoff indicates that it is a widely held philosophical position as well. His *sui generis* account, however, considers that intuitions are like perceptual experiences – pre-doxastic – that 'represent abstract matters as being a certain way. The idea is that in having an intuition, it seems to you that abstract matters are a certain way.'¹²² What Chudnoff indicates by his label of 'perceptual experiences', further, should be understood along the lines of intellectual perceptions, not purely sensorial ones. He summarizes his concept this way:

¹¹⁹ Eugene Sadler-Smith and Erella Shefy, 'The intuitive executive: Understanding and applying "gut feel" in decision-making', *Academy of Management Executive*, 18:4 (2004), 76-91.

¹²⁰ Elijah Chudnoff, *Intuition* (Oxford: Oxford University Press, 2013).

¹²¹ *ibid.*, p. 25.

¹²² *ibid.*, p. 25.

Let us introduce the notion of intuitive reasoning. This is reasoning that is based on intuition. It can involve memory, testimony, deduction, induction, and abduction. What distinguishes intuitive reasoning is that whatever beliefs are epistemically based on it are epistemically based at least in part on intuition. Notice that a bit of reasoning can be both intuitive and empirical. That is, it might be based on both intuition and sensory perception.¹²³

This definition does not, so far, appear to be saying much. What it does indicate however is that whatever knowledge an intuition may impart that would precede any judgment which follows from that knowledge, if indeed a judgment does so follow. On the face of it I do not see why this view could not be part of the psychological model we have so far considered, but Chudnoff's reasons for marking off his territory will become clear in the below, and some replies to them will also be offered at those points.

Throughout his book Chudnoff focuses on mathematical intuitions, perhaps believing that by using such examples his arguments will be most clear as he states that he considers intuitive insights to fall within the domains of mathematics, metaphysics, and morality. Thus in this way the abstract is kept central, and he thinks that his disagreement with the doxastic view can be found in the variance that, 'Doxasticists deny that we have, in addition to conscious judgments about abstract matters, pre-doxastic experiences that represent abstract matters as being a certain way... *Sui generis* recognize distinctive pre-doxastic intuition experiences that epistemically ground our beliefs about abstract matters.'¹²⁴ If the doxastic position is taken to involve automatic judgments of a 'pre-theoretical or spontaneous' nature (read: preconscious (/subconscious)) that Chudnoff

¹²³ *ibid.*, p. 16.

¹²⁴ *ibid.*, p. 26.

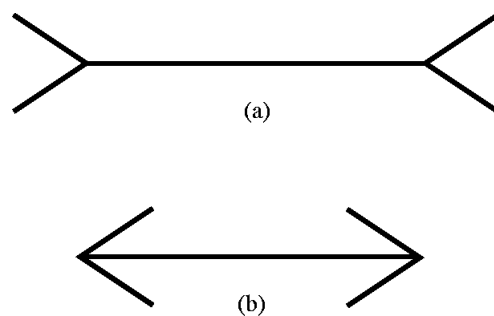
has himself characterized them as possessing¹²⁵ then what we might have here may really be just a slight confusion, or – on a possibly more charitable reading – a hint of a presumed background idealism of some manner. On the psychologically-based understanding of intuitions covered heretofore there is no ‘conscious’ (i.e. ‘aware’, on our terms) judgment involved *in intuitions* – whether about abstract matters or otherwise –, there are only *preconscious* judgments, and if such judgments contain purely epistemic features that can be separated off from the ascertainties themselves, or if they do not contain such separable features, is another issue from the judgments in question. The important matter, for our purposes at least, is that these judgments are automatic and generate behavioral expressions. Chudnoff, however, is not concerned with our point, his is a theoretical argument, but we do no damage whatsoever to our position that intuitions are mechanisms for decisions and/or evaluations (judgments) even if we grant that there may be an epistemic element that is even more automatic than or that precedes or supports the automatic judgments (maybe something like a Platonic Form?). What we cannot grant, on the other hand, is that aware (‘conscious’) decisions are *involved in* intuitions; at least not if we are indicating the root level of an intuition and that level only, i.e. if we mean an intuition taken on its own rather than as part of a broader process. Decisions incorporating awareness may be included at the (procedurally later) level of reasoning, however, and if we accept ‘intuitive reasoning’ as equating ‘reasoning based on intuition’ as Chudnoff has it (see the long quote above)¹²⁶ then we may really be speaking at a crossroads: on our account we wish to pull out the pure and undiluted intuitive core whereas Chudnoff takes that core plus what follows it and starts his work from that summation. To begin from this later point, as Chudnoff might be doing (on one reading of him), would on my view be a mistake however as I think it glosses important procedural matters. Reasoning – as

¹²⁵ *ibid.*, p. 25; Chudnoff does say ‘perhaps pre-theoretical or spontaneous’ but the psychological literature reviewed clearly has them as being that way.

¹²⁶ *ibid.*, p. 16.

‘reasoning’, as purposive and reflective thought – based on an intuition involves the aware and objective-driven analysis of the elements involved, at least one of which will contain information received via intuition, yes, but – and here for me is the key – the information so received took place at the *preconscious level* and therefore a judgment arrived at directly from an intuition (with its automatic and unthinking behavioral output) is a different matter from a judgment that has been arrived at indirectly – including what might be many other elements – with input also received from an intuition. ‘Intuitive reasoning’ would be an example of the latter, the indirect, while a *prima facie* reaction would be one of the former, the direct. If we think back to the two-level mental structuring proposed in Chapter 1,¹²⁷ the intuition occurs on the first, core level, while any reasoning would take place on the second, higher (or following, or possibly derivative) level.

That important variance highlighted, let us nevertheless continue with this issue of knowledge, for there is more to be found here. Chudnoff invokes the famous Müller-Lyer lines (illustrated below) to argue that we know them to be the same and therefore do not judge one to be longer even though it is perceived as longer.¹²⁸



I think this conclusion of Chudnoff’s confuses the position that our knowledge holds

¹²⁷ See especially the discussion in Section 1.1.C (‘Four accounts of the self: The soft realist position’).

¹²⁸ *ibid.*, see the discussion on pp. 27-28; image taken from: <https://www.pinterest.com/pin/418342252857562657/>.

within our (string of) judgments about such a case, and I believe expounding on this will be instructive and illustrative for our purposes. When I look at the lines I am very aware of judging line (a) to be longer than line (b) despite knowing that they are the same length. The illusion works because line (a) is not only perceived but judged to be longer – this is the automatic and preconscious evaluation that one cannot but arrive at. Knowing that the image is an illusion however allows me to rationally override that determination and remind myself that the two lines are the same length: the two-tiered process we saw above and just remarked on, the model we invoked when considering Kristjánsson’s ‘rationally grounded emotion’ concept that we also determined failed to place intuition as prior to reasoning in the decision-making process.¹²⁹ Has Chudnoff made what we take to be the same ‘lumping together’ error as we thought Kristjánsson did? It appears so, and we might add further that there are actually two types of knowledge involved here. The first type is intuitive knowledge (on Chudnoff’s account the ‘perceptual experience’ or intellectual perception that precedes my judgment (again perhaps he means to indicate a Form of some kind)) that line (a) is longer, and then the second type is knowledge that the lines are an example of an illusion and that their length is the same. If we accept Chudnoff’s argument we will have to say that our intuitive *knowledge* is mistaken in this case. If we do not differentiate between the knowledge element and the judgment element of intuitions (if they are in fact different elements) then we will merely say that our intuitive *judgment* is mistaken. Chudnoff appears not to want to admit either, and so argues that we do not judge (a) to be longer than (b) even though we perceive it that way. This seems to me to be a rather convoluted manner of expressing the familiar phenomenology involved when considering this illusion. The far simpler way, I think, to describe what happens is that we intuitively judge (a) to be longer than (b) and then in awareness (consciously) think better of it based on other, outside, knowledge. That, I would like to highlight, is precisely how

¹²⁹ See again Chapter 1, Section 1.1.C, ‘The soft realist position’.

the account of intuitions as judgments has it, and that intuitions can sometimes be wrong does no harm to that account. We should note though that all this does not mean that Chudnoff's account has nothing to add to our deliberations on intuition. Regarding matters of abstract knowledge in particular it may be useful, and perhaps even more apposite, to try and distinguish the epistemic and evaluative elements involved as distinct – and maybe even as garnered via access to an external ideal of some kind (a point on which we must remain agnostic for it would take us too far afield) – yet behaviorally what remains most salient in regards to the self is that intuitions function as automatic and preconscious decisions or evaluations, and this is so whether the knowledge involved precedes or accompanies the judgment itself.

As just demonstrated, and as accepted from everyday experience, intuitions are not infallible, and it is in the area of knowledge correction that Chudnoff's *sui generis* view joins with what he calls the doxastic position. In contemplating what he calls the typical intuition that no curve can be everywhere continuous and nowhere differentiable (as mentioned, Chudnoff's approach is primarily mathematical), Chudnoff demonstrates how reflection on the limit of an infinite sequence curve can lead to a new intuition: 'Error – specifically error at the level of the content of intuition – cannot only be avoided through further reflection, but also positively corrected so that one has an intuition experience that is accurate about the existence of everywhere continuous and nowhere differentiable curves.'¹³⁰ This matches one of the ways that the overall picture we have seen so far – of which Haidt's model is a good example – describes how intuitions can be changed, namely, via personal reflection. (Social influences may play a role here as well as presumably most of us need to be taught at least something about the types of curves involved, infinite sequences, or both.) Even granting Chudnoff the *sui generis*/doxastic polarity he wishes to

¹³⁰ *ibid.*, pp. 102-103.

establish, the fact that both purported views of intuition accept error correction in this way does lend credence to the more standard account that has been heretofore discussed. Indeed, it may be that in seeking to establish a sharp line between his view and that of others Chudnoff exaggerates differences that are merely slight or not really there at all. Regarding beliefs and intuitive justification he writes that:

if intuitions about reasons for action play action-guiding roles, their ability to do so does not depend on their leading to beliefs. You can act in light of the justification you have for believing that you have a reason to act without forming the belief that you have that reason to act. In fact, this is likely the norm. Beliefs are mental states with a certain degree of permanency. Most of our actions have little significance beyond the moment of their occurrence. It would be silly to form standing beliefs about what reasons you have for all the actions you perform in life.¹³¹

In essence, Chudnoff is here stating that intuitions justify and therefore guide actions more than beliefs do, a point that I think very few, if any, of the people holding positions he terms doxastic would disagree with. That justification is involved does imply awareness-based decision-making, however, which appears to be at variance with intuitive judgments as preconscious decisions or evaluations, but that is only on the surface. If the picture of intuitions as preconscious judgments is accepted then they will give rise to both automatic behavior and reflected-upon behavior. One might ‘just know’ something and use that in one’s calculations for action (when employed) the same way that one might unthinkingly take the identical something into (preconscious) account and react, with no further calculations made. Certainly such an interpretation would appear to match with both Chudnoff’s view and with the picture that has otherwise emerged so far.

¹³¹ *ibid.*, p. 169.

Yet one aspect remains unclear: How exactly does knowledge fit in? Chudnoff has given us an account where abstract knowledge precedes judgment and focuses on mathematical insights to ground his case. What of the everyday intuitions like my wife's about our apartment? Perhaps Sadler-Smith and Shefy's offer of intuitions as a composite of knowing and sensing¹³² is the simplest way to think of them. Intuitions can deliver knowledge and they can deliver judgments. Often they will deliver both – there need not be an impasse between these accounts. Whether their content connects us in some way with what may be ideal or transcendental – Chudnoff does explicitly suggest intuitions exhibit some kind of world-to-mind relationship in addition to the mind-to-world one we usually associate with knowledge¹³³ – or whether our preconscious and its vast capacity merely processes information and offers up analyses as intuitions I cannot say; for the sake of prudence and due to our study's limited scope I wish to maintain the agnosticism mentioned above, much as I find Chudnoff's notion on this point appealing. Nevertheless, what is applicable for our purposes in exploring the self is that the model of intuitions as preconscious judgments that give rise to behavior has been shown to be robust. What remains now is to sort out the role emotions play in all of this.

When we reflect on our emotions the noticeably stronger experiences tend to come to mind first, such as anger, fear, joy, depression, et cetera. Usually in hindsight we can point to roughly when these episodes started and when they stopped. We describe feeling a certain way at a certain time and we understand that the decisions we make during those periods will be colored by the ways that we are feeling. Similarly, as the above has attempted to show, intuitive reactions will often affect the way we think and behave without our

¹³² Sadler-Smith and Shefy, *op. cit.*

¹³³ Chudnoff, *op. cit.*, see the discussion on p. 158ff, especially pp. 161-165.

necessarily being aware of such taking place. Given that in the heat of the moment, so to speak, we also do not pause to consider or even notice the feelings we are experiencing at that precise instant, how do the two differ? Are not affective influences and intuitive influences more or less the same thing as far as the end behavioral product goes?

On the surface it might seem that way, and this appearance may account for much of the mixing between the two in the literature, as was demonstrated by Kristjánsson's account and as will be shown through other examples below. There are important differences between them however, and, I believe, there exists also a procedural difference in degree. One initial area where upon consideration the variance between an emotion and an intuition becomes clear is in the perhaps surprising fact that the former is often actively taught. Peter Goldie points out that, 'The process of teaching a child how to identify things which are dangerous is typically one and the same process as teaching that child when fear is merited.'¹³⁴ Goldie has it that through such learning we come to understand the evaluative properties that warrant a certain type of response, and these responses will include aspects of thought, feeling, and action.¹³⁵ It will be noted that this idea of a 'response' (not defined in detail), coupled with such diverse areas as thought, feeling, and action, approximates the picture we have seen of how intuitions operate.¹³⁶ While this is an interesting crossover, what importantly needs to be recognized here is that the emotional responses under discussion are not *judgments*, they are *replies* to perceived evaluative properties; the distinction is subtle but crucial. To elaborate: A judgment is an evaluation, a response/reply will stem *out of* such an evaluation. There is the processing and then there is the output. An emotion is part of the manifested behavior generated in an organism

¹³⁴ Peter Goldie, *The Emotions: A Philosophical Exploration* (Oxford: Oxford University Press, 2000), p. 30.

¹³⁵ *ibid.*

¹³⁶ Although Goldie's focus on coming to understand does not itself fit as that is not conceptually a part of intuition, or at least not directly.

following the brain's data analysis of perceptions and/or stimuli; or at least that is the case I will attempt to make in the following. When we feel fear we certainly do tend to associate a decision with it (e.g. fight/flight), but I want to argue that that decision is a separate process from the feeling of the fear emotion and that it indeed precedes the fear emotion. If you will, allow me to expound using Goldie's account both as an informative basis and as a foil against which to contend.

Goldie explains further what he means in casting emotions as 'responses' with the following:

For each sort of emotional experience there will be a paradigmatic narrative structure – paradigmatic recognitional thoughts, and paradigmatic responses involving motivational thoughts and feelings, as well as bodily changes, expressive activity, and so forth – and this will be epistemologically central to the concept of that emotion and to how we learn to apply it.¹³⁷

This helps explain why the distinction between our affective responses and our intuitive reactions appears so nuanced; we must take care not to equate paradigmatic recognitional thoughts involving appropriate (and proportionate) emotional responses based on our emotional education with intuitive judgments based on such paradigmatic thoughts and responses but occurring spontaneously as part of our biological functioning. Although both might indeed be inbuilt (biological functioning of course, but as will be put forward there are core root emotions as well as those emotions which we 'learn' as we grow), they are nevertheless layered. What is difficult to appreciate as a normal adult is that although one's emotional responses seem just as automatic as intuitive flashes strike us as being, in many

¹³⁷ Goldie, *ibid.*, p. 33.

cases we have been taught (explicitly or implicitly) to have those emotional responses and it is only via long experience that they become automatic in the way they do. My toddler and newborn (at the time of writing) daughters have shown me much about this process as they react emotionally to us their parents in ways that are natural but not always spontaneous. In the case of the toddler, she really does not know what about a given situation warrants joy or grief, but she has developed enough – unlike her younger sister – to realize that some reaction is warranted (though the descriptive ‘some’ can be (tryingly) an understatement of the strength of her reactions). She observes us and responds as similarly as she is capable, but to what degree she conceptually grasps the underlying cause and effect relationship can only be guessed at; certainly her understanding is being built day-by-day – but it is having to be built.

The situation though is unfortunately not even as clear as that for one can – and frequently does – experience emotions which do not seem to have been based on any education whatsoever, the fear of snakes being an example of such a rudimentary emotion that is often cited (i.e. what I referred to as a ‘core root emotion’ immediately above). How should we think of such? Goldie describes interpreting emotions in this way:

What often happens, I think, is that we first have an emotional response towards an object, a feeling which is often quite primitive... Then, in self-interpretation, when we become reflectively aware of this feeling towards the object of the emotion...we also normally seek to make it intelligible by looking for the identificatory and explanatory beliefs [involved]... What really comes first is the emotional response itself – the feeling of fear towards the snake – and not the

thought that its bite is poisonous and the thought that poison would harm me.¹³⁸

Surely we can think of the initial response in such a case as automatic and preconscious? Strikingly, this description fits almost exactly the manner in which Haidt outlines what occurs when we have an intuitive reaction and then seek to interpret it.¹³⁹ Where can the dividing line between these kinds of ‘primitive’ or ‘core’ emotional responses and intuitive reactions be placed? Does not each appear to be a type of judgment? Evaluative? It does seem that way, but I think that on further analysis a means will present itself, and so we look for our answer by going deeper into what Goldie has to say about these particular types of emotional responses.

Following his snake illustration Goldie discusses the situational fear involved where one is likely to be fired (involving beliefs and desires), and where one is likely to be hit by a bus; Goldie’s point in the passage is that beliefs need not play a causal role in explaining emotional responses, but he makes no mention of intuition being involved in the decisions and behaviors taken, instead calling them examples of complex and primitive psychological episodes.¹⁴⁰ In reflecting on what might be involved here, while I do think that it is possible to set aside the former firing scenario by simply stating that although intuition may or may not have a significant role to play in the considerations involved where one fears losing one’s job, it would appear that given the picture of the procedural elements we have encountered repeatedly in our considerations of the psychological evidence intuition would certainly have a role to play in jumping out of the way of the oncoming bus. If there is any scenario that calls for rapid, nonaware data processing and an evaluation followed by a behavior it would be this, or something very similar to it. One

¹³⁸ *ibid.*, p. 45.

¹³⁹ Cf. Haidt 2001, *op. cit.*

¹⁴⁰ Goldie, *op. cit.*

would not even need to think (rationally, that is) about such a reaction, it would be an automatic and preconscious decision immediately generating an action; all the hallmarks of an intuition that we have seen so far (and it would moreover include the (intuitive) knowledge element of what could happen if one were so struck). One would likely also feel fear, of course, but that Goldie ignores intuition altogether at least implies that he might be overlooking them and thus making the same mistake that Kristjánsson did in his account where emotions were merged with intuitions rather than being considered separately (although since Goldie wrote before much of the subsequent psychological research on intuitions was released he may simply not have been cognizant of such considerations).

When we work through cases like these they do appear to strengthen our conclusion that intuitions are automatic and foundationally rooted judgments, and that while such judgments will typically co-occur with emotional responses the processes involved are distinctively generated. Returning to the situation where one fears being fired we can find the same: thinking that one might be let go, if it is not simply based on an aberration such as a paranoid condition, it would stem from an evaluation, a judgment, regarding one's overall working conditions. This assessment could well be the result of an aware and rational reflective analysis, and having arrived at it one might then feel fear if one considers one's conclusion to be accurate and/or likely. This may or may not include an intuition somewhere along the lines. If we accept the two-tiered mental model described earlier though it strikes me as at least being a possibility that the rational thought that led to the worry of being fired and the feeling of fear might have been triggered by an initial intuition. Perhaps one's boss reacted in an unusual way that was previously observed, or a bit of work took an unpleasant turn, or a headline alluding to widespread economic pressure and downsizing was noticed while scanning a newspaper. There are any number

of places where intuition might play a role, but whether it does or does not we are nevertheless able to understand it as separable – and indeed as separate – from emotion. Well and good, but we still need to see on what grounds we might think that intuition does come first within our layered mental model, that it is even more primary than an emotion in those instances where both occur. Our exploration continues.

From these considerations we can understand that emotions and intuitions often go hand in hand and on the surface may even seem to be one and the same. This is particularly so in regards to the functions they carry, which will frequently intersect and overlap. We have even so asserted that the two ought to be considered as discrete, and that one precedes the other. This, though, has likely not been satisfactorily established for all readers and so let us now turn to a distinction that Goldie does make – one between emotions and feelings (which he also calls ‘emotional episodes’) – in order to try and establish a more detailed view of what is involved in an emotional state so that we can thereafter see if such will help us find its difference from an intuitive state. Goldie first explains the need to distinguish between reflective (aware of while occurring) feelings and unreflective ones, and having done that he turns to argue that feelings need not pertain during an entire emotional duration:

But once the distinction between an emotion and an emotional episode is clear, I do not need to insist, for example, that my enduring love for her (the emotion) need involve feelings at all times, even when I am playing football, although it will be true that my emotion at that time involves various dispositions to have loving thoughts towards her.¹⁴¹

¹⁴¹ *ibid.*, p. 69.

The view as presented here is structured with the emotion in the background, as it were, and out of which emotional episodes/feelings are then generated as instances of that emotion and/or dispositions based in that emotion. Maintaining this, a few pages later Goldie discusses the desire for something as an indirect or second-order emotion or feeling in the same way – as part of the background enduring emotional state¹⁴² – and this conceptual viewpoint, I think, presents an interesting analogy for the manner in which emotional states (on this understanding of them) may inform intuitive judgments. Could it be that an ongoing emotional condition underlies even intuitions and forms the deepest section of our procedural model? Let us consider the terms, structure, and implications of this perspective because if we were to accept it that would imply a tri-factored approach to the relationship between emotions and intuitions and shift the understanding we have carefully built up. Adopting Goldie’s stance would mean that we would have the enduring emotional state as the ground out of which both an intuitive judgment and an emotional episode/feeling would spring, yet we would still not be clear about how the latter two are related to one another, both in regards to timing and to interaction. Moreover, another difficulty can be recognized in an additional sample that Goldie presents: desire. Although he does recognize the common view of desire as pertaining to realization or satisfaction in his discussion of it, he nevertheless focuses solely on desire in the sense of ‘desire for’ (i.e. possession), and not on ‘desire of’ (i.e. realization or satisfaction).¹⁴³ Failing to do this is instructive in the consequences it carries for his overall scheme: Would not the ‘desire of’ sense count as the kind of long-term background emotion he distinguishes from a short-term feeling and therefore make it no longer indirect or second-order? If so, then how could the mere addition of one preposition or another change desire from being a second-order to a first-order emotion? Is this merely a deficiency of the English language

¹⁴² *ibid.*, discussed on pp. 78-81 as part of his ‘feelings towards’ concept; an idea which, while interesting, is inapplicable here.

¹⁴³ *ibid.*

that fails to account for the conceptual differences between these types of desire by inadequately employing the same term for both? Or are we perhaps asking too much since desire as ‘possession’ could also arguably be considered ‘realization or satisfaction’? Most crucially, would a first-order emotion have to count as an emotional state or would it be a feeling? Why not just call an enduring emotional state a mood or a trait?

To answer this objection Goldie might direct us towards his discussion of emotions and moods in which he defines emotions as being specific and moods as being non-specific, though he does also caution that on his view emotions should not be seen as fixed and impervious to feedback via actions born from the emotion.¹⁴⁴ Although Goldie earlier made the case for emotions being durable (versus emotional episodes/feelings), that they can in fact last for years at a time, here he seems to perhaps place moods as even more fundamental (equally fundamental?) than they, stating that just as emotions can come out of moods (e.g. anger out of irritability) they can also settle into moods, and moods can consolidate into traits (e.g. resentment into a resentful person).¹⁴⁵ This approach appears to be building some kind of relational pyramid, but one in which the levels can interact and trade out with one another, yet there is also a sense in which the picture remains unclear. On this view, for instance, it may be that ‘mood’ is being used too vaguely to really get a clear understanding of just what element does what.

In reflecting on these issues that have arisen as we worked out the details of Goldie’s account I believe on the whole we can abandon his emotions/moods aspect without suffering any theoretical loss and here is why: Goldie bases his conceptual construct on the musings of the Ulrich character in Robert Musil’s *The Man Without Qualities*,¹⁴⁶ thoughts

¹⁴⁴ *ibid.*, see pp. 143-145.

¹⁴⁵ *ibid.*, see p. 150.

¹⁴⁶ Robert Musil, *The Man Without Qualities*, trans. by Sophie Wilkins and Burton Pike (New York: Alfred

presented as referring to ‘specific emotions’ and ‘nonspecific emotions’. Musil, however, presumably uses ‘emotion’ in that work the same way the rest of us do: in its everyday sense of a feeling (which in Goldie’s hands becomes an ‘emotional episode/feeling’ as distinct from the enduring ‘emotion’); this would mean that Musil does not make the long-term emotion versus short-term feeling separation that Goldie does, he rather equates the two as is done in common speech (e.g. having an emotion = having a feeling). If so, then Musil’s usage of ‘specific emotion’ in his novel that Goldie picks up on would simply indicate a feeling directed towards a certain something, and his ‘nonspecific emotion’ would point to a feeling directed towards nothing in particular – which in other words is what we call a ‘mood’. The distinction of ‘specific’ and ‘nonspecific’ is therefore between an episode and a mood, and hence it seems that we could substitute ‘emotion’ in Goldie’s sense of it as an enduring emotional state for ‘mood’. This though is the very distinction drawn between ‘emotion’ and ‘emotional episode/feeling’ by Goldie’s earlier argument, and therefore in this overlap either ‘mood’ is really nothing separable on Goldie’s account or all that he established in arguing for an ‘emotion’ as an enduring state was to exchange that term for how the equal (and equal even on Goldie’s definition once we have gone through and worked out the preceding) ‘mood’ is typically used (or, perhaps in the case of something like love, ‘character trait’).

If we can consider enduring emotional states as being moods and as contributory to the expression of traits then it becomes clearer how such form the background out of which emotions (i.e. in the regular sense of ‘feelings’) arise out of either any given mood and/or trait, and we can perhaps propose too that intuitions may arise – and arise independently – also from traits (recalling that traits denote the results of genetics, upbringing, experiences, et cetera). Let us deliberate on both; intuitions first. We have in the above seen that

A. Knopf, 1995); see Goldie, *op. cit.*, long sections of Musil’s work are cited on pp. 144-147.

intuitions are automatic and pre- or non-reflective judgments, and that they often – but not always – contain knowledge. We have seen too that intuitions, despite their fundamental status within our two-tiered mental model, can nevertheless be purposefully shifted, trained, strengthened, or weakened through internal efforts and via external (social) influences. Here in these strivings the relationship with traits might be clearest: I want to give myself entirely different or somewhat altered intuitions because I want to act otherwise than the way I notice myself acting (and, more to the point really, *reacting*), thus I attempt to achieve internalization of a desired trait like patience so that I may become a naturally (intuitively, automatically) more patient person. In this way, as remarked, we also find Kristjánsson's moral self with its Aristotelean type training, a view which was not determined wrong in its treatment of intuitions in this regard so much as incomplete.

We have also understood that emotions are in some ways similarly built-in responses to the world, but in many other ways are educated responses. The former (built-in) type may be considered this way: I see a snake, my intuition tells me to avoid it (evaluative judgment), and I feel fear (response – but I may not know why); on the other hand, the latter (educated) type may be viewed this way: I see an oncoming bus, my intuition tells me to avoid it (evaluative judgment), and I feel fear (response – and I do know why: because I have learned that it would be terribly harmful to be struck by a bus). In both cases the intuitive judgment is foundational but, as outlined above, could be approached through the purposeful working on of one's personal traits (and again, traits result from genetics, upbringing, experiences, et cetera, they are neither intuitions nor emotions although all are interactive and connected within the brain's processing systems). Emotional responses may similarly be trained – we can even try to attain moods as many of us will know from our own personal endeavors – but that they can does not detract from their differences with intuitions, nor does it imply that they necessarily precede (or admittedly, and just from this,

that they must follow) intuitions. Rather it implies that the emotions one has will help to augment one's intuitive evaluative judgments: emotional responses either reinforcing intuitive reactions or emotional responses being at odds with intuitive reactions and thereby contributing to the gradual shifting of those intuitive reactions. (Another thing my daughters have taught me, through a changed emotional response, was to get rid of my intuitive judgment of disgust/avoid regarding feces – apologies to delicate readers for the example.) We have now finally separated intuitions and emotions, and although we have not been able to definitively and beyond doubt establish that intuitions always and only come before emotions we have good reasons based on our cognitive model (which has proven to be dependable and resilient) for believing that in most to all cases they do – or at the very least that they occur contemporaneously (and not later) than emotions.

2.3 Bodily and embodied issues

We return now to our place of departure, that core which Quassim Cassam describes thusly, 'the point of talking about the self or subject is to identify a locus of thought, perception and action. Yet perception and action require embodiment, and thought is inseparable from perception and action.'¹⁴⁷ Body, flesh, fullness.

The idea of the self as a pure intellect, or as an aethereal pilot in a physical machine, or even as a brain in a vat, makes little visceral sense to us given the type of beings we are and living in the world we inhabit. When we think on such we are immediately reminded of Princess Elisabeth's objection to Descartes' system: How can an immaterial mind direct a material body?¹⁴⁸ Surely the self must be embodied in some way? Indeed, whatever 'we' are our bodies are clearly very important to how we consider ourselves in a typical,

¹⁴⁷ Quassim Cassam, 'The Embodied Self' in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), pp. 139-156 (p. 154).

¹⁴⁸ Hatfield, *op. cit.*

unanalyzed (e.g. pre-Cartesian doubt regiment), everyday fashion. Psychological studies moreover indicate that our sense of our bodies, of – in Philippe Rochat’s phrasing – our ‘own feeling bod[ies] as a differentiated entity among other entities’ appears to take root in our referential thinking from a time shortly after birth; as infants we understand ourselves as being embodied, perceived, and spatiotemporally situated.¹⁴⁹ Initially this understanding is that of an ‘ecological self’ (or ‘minimal self’ or ‘minimal subject’), not yet filled out by the broad emotive sets that Kristjánsson describes, not yet a ‘me’ in a fuller sense.¹⁵⁰ The wording I used above is intentional: ‘we understand ourselves as being embodied’; we remain unable to get past the notion that ‘we’ are something inside our bodies looking out, for the inner experience of thought seems to cloud all else even while it establishes the view that one is somehow significantly body bound. How then does the natural comprehension of being an embodied creature play into the wider understanding that we have of ourselves as selves? In the affective and (soft) realist self concept that we have been discussing, what is the relationship between the self (our adjusted versions of Kristjánsson’s Set 1 + Sets 2 & 3 + consciousness + the presence of a body) and the body proper, as it were? Can the two be separated (can we ‘have’ a body), or must they be one and the same (we ‘are’ a body)?

Our opening comments appear to have already answered the first part of that last question in the negative: a disembodied self strikes us as nonsense, if we ‘are’ something like a Cartesian pure intellect then what are we? Some type of immaterial aether? If so, how do ‘we’ interact with ‘our’ physical bodies? (In Descartes’ account the body functions through the movement of ‘animal spirits’ which pass through the pineal gland and it is there, within

¹⁴⁹ Philippe Rochat, ‘What is it Like to Be a Newborn?’ in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), pp. 57-79 (p. 64).

¹⁵⁰ *ibid.* Some thinkers stop here with the idea of a self only ever needing to be a minimal self, an issue to be discussed in the next chapter as a potential objection to the self concept being advanced in this study.

that important gland, where the purely immaterial mind is able to interact with them.¹⁵¹) What would it mean for a self merely to have the presence of a physical body but not be embodied, as might appear possible from the definition of the self we have established? If we rather ‘are’ material in the sense of the self being equated with the functioning of the brain (perhaps by claiming the self as being ‘mind’ – a ‘living brain’ – as opposed to simply understanding ‘mind’ as ‘brain’, that is, the functioning organ versus its compositional matter) then could a floating brain perceive and act, thereby generating the necessary thought and its nexus that Cassam described? If so, would that be sufficient for a self? We have claimed that the commonsense view of a substantive self seems to have worth, yet what kind of substance are we talking about? The case is far from being a simple one and thus some further discussion is warranted; in it we will take examples both from the realms of experimental evidence and from thought experiment deliberations.

To begin with, consider the case of the Rubber Hand Illusion, in which subjects view a rubber hand that is being stroked by a finger in the same manner in which their unseen hand is being simultaneously stroked. Subjects in this condition who experience the illusion see the rubber hand as more like their own hand than subjects who do not have the illusion; in other words, the false projection of ownership actually alters perception for the affected subjects. That is somewhat remarkable, and the biological effects are even more surprising. For those subjects, the ones under the illusion, the rubber hand is quite literally internally experienced as being their own hand, its location is thought to be the location of their ‘real’ hand, and the skin temperature of their in fact real hand decreases during the illusion (with this decrease corollary to the strength of the illusion), indicating that a disruption of cognizance of what is (a part of) the body may lead to a disruption of the

¹⁵¹ René Descartes, *A Discourse on the Method*, trans. and intro. by Ian Maclean (Oxford: Oxford University Press, 2006).

functioning of normal physiological regulation.¹⁵² What might we conclude from such astonishing evidence? At first blush, we might be tempted to think that if the brain can be deceived in this way – and the whole physiological system responding with it – then it appears that whatever our selves are, if they are brain-based, they must be separable from the physical body that they are connected to. Thus both the rubber hand and the real hand are not a part of the ‘self’, which is why subjects were able to experience them as interchangeable: without a secure connection anchoring this hand that types to ‘me’ it could be cut loose while ‘I’ continue on quite as normal. Results that studies like this one reveal are indeed interpretable as being suggestive of such a disconnection. Would this, we may query further, be akin to what Descartes arrived at via his program of radical doubt? There the meditator knew that he existed but that, at least in the *cogito* argument’s initial conclusions, he had neither body nor senses.¹⁵³ Might such thoughts find scientific, empirical support here?

Lest we get carried away a pause is called for at this juncture, for in the preceding sections of our study we have already – for many varied reasons – rejected a fully disembodied self, and if instead we now, based on this illusion, claim an exclusively brain-contained self then we must remember that the brain is no less a part of the physical body than the hand, even if its functioning and purposes are very different. Yet there does seem to be something to this, and the brain is after all a very special part of the body: if then the brain, being the regulatory control center that it is, can be misled into altering standard physiology on the basis of an illusion that effectively replaces a false hand for the one actually connected to an arm then perhaps the brain is distinct enough from the rest of the body to warrant a differing status. What would that status be? It would appear to have to be as the housing of

¹⁵² Manos Tsakiris, ‘The Sense of Body Ownership’ in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), pp. 180-203.

¹⁵³ Hatfield, *op. cit.*

the self (within the current context; there are of course other ways lines could be drawn, e.g. cellular, nervous, biological role, et cetera), whose view of which hand belongs to it determines the corresponding perception and regulation, and in turn indicating a further claim that the self is embodied in the brain, that it views the rest of the body from that perch (as it were). This is little different in practice from a disembodied self, but it is different enough to shift the self out of the abstract and into the concrete. The self, on this view, would either be an inhabitant of the living brain (as above) or a product of it, although whether this means a product solely of the living brain or instead a product of the living brain as it functions from within a whole body remains unclear. If the former is the case then the self is a ‘something else’ that takes up residence in the brain, and we find that we have come back to a Cartesian concept and may as well call the self a ‘mind’ or a ‘soul’. If, however, the latter is the case then we might have something of a breakthrough on our hands (pardon the pun).

We recall that in the foregoing we tried to establish an argument for taking the self as real in the way that consciousness is considered to be real – by being emergent –; might an idea like this offer any clue to what may be happening as we follow this lead? It does, or could, in that it tells us that the self is not located *in* the body, nor *is it* the body, it is rather *of* the body, specifically of the brain; at least that is what we can suggest at this early point in our bodily investigation. The self – if we are onto something here – is neither contained nor connected, but rather generated. Applying all of our study’s preceding to this argumentative advance, we would have the self as something that develops over time, yet not – given the various ways the self shifts and can be shifted – in a linear manner that builds towards a completed whole, instead in an amorphous way: filling in details, deleting some elements and adding others, maintaining core features for long – perhaps very long – stretches, yet altering nuances and expression all the while. Such is the unfolding and

interacting of Kristjánsson's three self-composing sets adjusted in the ways we described and as outcomes of the brain's standard operation. The self, by this view, would be something beyond the body that nevertheless needs the body to exist. It would be an 'embodied' self engendered by the normal functioning of the brain. We – as a self – 'have' the body that we 'are' at a minimum as far as the brain is concerned, but we suspect that based on the many aspects we have hitherto explored other features of the body will prove necessary as well, particularly when we recall the formative and feedback issues connected with social input and embeddedness.¹⁵⁴ We therefore continue.

To test this suspicion about the whole body's role in the self – not just (potentially) the brain's –, let us unpack what we are dealing with using some counterfactuals to aid our thinking. Imagine first a case wherein a brain has been successfully transferred to a vat: what would remain of the self, if anything, in such a scenario? Would that brain still have its self-constituting, self-comparative, and self-conscious emotions? What about the other elements of our defined self such as consciousness? Would the brain itself count as enough of a bodily presence to constitute the bodily presence qualifier that our definition also contains? Cassam's opening quote to this section notes that 'perception and action require embodiment, and thought is inseparable from perception and action',¹⁵⁵ causing us to consider that thought, at least, would not be possible in such a minimized case as the brain would lack a way to perceive and to act (unless one counts thought itself as action, but then if thought is 'inseparable from perception and action' both elements would at any rate be necessary and thereby thought-as-action (combining the two) insufficient for it would still lack perception). We might yet, however, consider the brain to be a satisfactory bodily presence for a self since its functioning as an organ is what we determined to be the

¹⁵⁴ On those see especially Section 2.2.D above, 'Selves, personal identity, and whole persons'.

¹⁵⁵ Cassam, *op. cit.*, p. 154.

generative source of the self, and thus also we might consider the physical brain as the location of the self's necessary embodiment. We may grant too that consciousness is maintained given the stance expressed that consciousness can be considered present if at least one part of the overall consciousness network is working (that network being composed of the brain's many domain-specific consciousnesses that are typically communicatively linked together, as described). Yet even accepting all these points we are still unclear on the importance of the absence of thought just alluded to (without perception and action), and we are also unsure of how emotion might fit in.

As our claim has been that Kristjánsson's three emotive sets – adjusted in the manners described – are central to the self we might wish to consider emotions first, keeping in mind that although thoughts often contain emotional elements, are emotionally directed, or are in some way about emotions, they are not strictly speaking emotions, nor even necessarily emotion related. What would it therefore mean for this envatted brain to have emotions? The brain would be incapable of acting on any of the chemical affective influences it was experiencing (and without any stimulants in its environment and lacking perception we must assume such would be minimal), incapable of forming thoughts from them (or from anything, again lacking perception and action), and therefore incapable of sustaining any of Kristjánsson's three sets. Moreover, what kind of affections might be generated for a brain in this condition? It is floating in a stable fluid, or anyway placed in a stable manner, fed what it needs to sustain it (by whatever method has been devised for the purpose), and cut off from any and all interactions with its world – even mere perception. That I naturally use the word 'it' to describe this brain, and that the reader has most likely not found the term objectionable, is informative: in fact the brain would still be either male or female depending on its original (i.e. pre-transfer to the vat) source as both testosterone and estrogen are formative chemicals regarding how a brain functions (the old 'men are

from Mars, women are from Venus' yarn),¹⁵⁶ yet stripped of the rest of the body we have no compunctions about disregarding its gender. The person's memories would probably remain, however, as these are considered to be physical structures consisting of differentiated neural tissue that release and are associated with context signals within networks (say, for a word and its meaning).¹⁵⁷ What role these memories might play for such a brain is unclear, but if the brain does retain consciousness in some way we may speculate that these memories could be accessible, though again not in a thought-directed manner, or at least not in a thought-directed manner that would be something we could recognize as such. Nevertheless, even granting consciousness and access to memories, barring the possibility of thoughts about them (as well as action in and feedback from the world) makes it difficult to see how an emotion in even a broad Goldiean background sense might come into play for the brain (à la mood: Goldie's 'emotion' as an enduring state versus his 'emotional episode/feeling'). Whatever does exist for this brain without emotions it clearly would not have a self in the sense we have been arguing for, nor even an ecological or minimal self as it lacks perception and spatiotemporal situation. (Moreover, in thinking on this it becomes clear how generous it is to consider only the brain as being a sufficient source for embodiment.) If, in its perceptionless and actionless way this brain could still be said to have experience, such would – at most – amount to accessing its own memories in some indescribable and inoperable way. On the other hand, if the brain in question were to be given a means of perceiving and acting within its world then we can imagine that it might start with an ecological self in the way a newborn does. Our next case therefore gives the brain a body.

¹⁵⁶ For a brief and very general overview see Robin Nixon, 'Matters of the Brain: Why Men and Women Are So Different', *Live Science*, 01 May 2012.

<<http://www.livescience.com/20011-brain-cognition-gender-differences.html>>.

¹⁵⁷ Lance Knowles, 'Mind blocks', *Language Magazine: Journal of Communication and Education*, 7 (2008), 28-34; on memory formation in the context of language acquisition see also Prahlad Gupta and Brian MacWhinney, 'Vocabulary acquisition and verbal short-term memory: Computational and neural bases', *Brain and Language*, 59 (1997), 267-333; and Birgit Henriksen, 'Three dimensions of vocabulary development', *Studies in Second Language Acquisition*, 21 (1999), 303-317.

Our sample brain, it can be the same brain as above, is now attached to a robotic body equipped with sensors that allow it to ‘see’, ‘hear’, ‘feel’, et cetera, its environment. For practical purposes we can keep the brain in the same type of sustaining vat, it need not be put into a ‘head’. The brain can now experience and interact with its surroundings: it has a ‘body’ and is aware of its being embodied, that it is perceived and perceiving, and that it is spatiotemporally situated, initially cognizant of all those same aspects that a newborn is¹⁵⁸ as she faces a world so startlingly different from the womb. In addition the brain, equipped and defined now to fit with the concerns raised by Cassam in its ability to perceive and act, can also think in the manner to which we are accustomed via the interplay of those capabilities, thus rounding out our picture of a creature not terribly unlike us. What kind of self might this cyborg have?

If the cyborg could be built with an actual newborn’s brain (and not simply with those aspects that are like a newborn) we might think that it starts out with the same kind of ecological self that a newborn does, but that case is less interesting than the stickier one we have set up where a brain is first transferred to a vat and, we suppose, the brain is an adult’s brain. In that case the already fully developed (from its pre-surgery life) and transferred brain would at the outset conceivably retain its self-constituting, self-comparative, and self-conscious emotions (Sets 1, 2, and 3; we can perhaps consider these as leftovers from the time immediately prior to the transfer procedure), and would moreover also of course have consciousness and bodily presence, thus containing all of the elements of our self as defined above. We can assume this to be the case if we imagine what it would entail for the process to unfold: a person would have to be put to sleep prior to their brain being removed from their body, and once so removed the brain would have to

¹⁵⁸ Rochat, *op. cit.*

be kept in some form of preservative stasis while it was enclosed in the vat that is attached to the robotic body. (Conversely, if the brain were first in the vat and then that vat attached to the body little would change as perception and spatiotemporal awareness would not occur until the brain had been hooked up to its ‘senses’.) For the person involved we can think of this as going to sleep and then waking up again, with no physical interference to the brain other than its spatial placement and subsequent nervous and muscular re-connection to the new sense organs. The brain’s crucial regions, such as the temporal lobe, frontal lobe, cerebral cortex, amygdala, thalamus, et cetera would all be untouched.

If we put ourselves in this situation we can immediately see how shocking such would be: lying down on the operating table, your last memory is of the gas mask being put over your face and a surgeon looking down at you, then suddenly you hear a beep, open your eyes – only they are not eyes, they are sensors and by ‘opening’ them they are simply turned on – and then raise an arm in groggy wonder to see that it is made of metal. Metal? You jolt upright, demand a mirror, your voice coming out of a place somewhere further down from where it should be and sounding all wrong in its artificiality. The surgeon cautiously passes over a small hand mirror and you lift it to be faced by an unblinking brain floating in a vat of viscous fluid and affixed to a metallic neck. You bend your ‘head’ down to point the sensors at your body and discover the full horror of what you have become. In looking at the case this way we see how almost immediately the contents of the three emotive sets would change, and likely change very drastically. Furthermore, although for dramatic purposes we have in our little vignette imagined that we the subject was not advised about what the results of the surgery would be, even if we had been and therefore fully expected to see our new metallic arm, torso, hear our new machine-like voice, et cetera, would that really alter the transformations to the self that would henceforth occur? Pre-knowledge might help with the degree of reshaping involved, but would that really matter in any final

analysis of this new me?

Let us try and break it down. While we might think that some parts – perhaps certain commitments – of the self-constituting ‘emotions’ that make up Set 1 would have the most durability (e.g. if a pre-brain transfer person is dedicated to helping ease the burden of poverty by regularly donating to a charity/charities, then this would not be likely to alter even given the radicality of the physical transformation undergone), such is not necessarily the case. Recall that Set 1 consists of items like commitments, but also traits, aspirations, and ideals, and as pointed out in the critique of Kristjánsson’s account these features are not really emotions.¹⁵⁹ If I wake up and find myself a cyborg I may keep some of my former commitments but I will likely change others, possibly many, and my aspirations and ideals would no doubt require much adjusting. I may also try to retain my traits, but as the following will show this too would be very difficult. As I adapt to my new body my self-comparative and self-conscious emotions (Sets 2 and 3) would naturally go through a severe process of revision as I recognize the very different reality I now inhabit, and perhaps this would have an even more profound effect on me than my internal considerations, particularly given what would occur on the external, social side of the emotional influences affecting the self (the self’s sets). Others would suddenly, and for them quite naturally, react to this new me in a manner shockingly distinct from the ways the same others used to interact with me before my cyborg rebirth. Amongst those who were especially close to the former me, relations with my cyborg being may eventually approximate what they were (if my family and friends are very generous, patient, and open-minded), but of course my new cyborg way of life, abilities, physical presence and appearance, expectations, goals, dreams, et cetera, would all be so disparate from my former human ones that the transformation would largely amount to hitting the reset button

¹⁵⁹ Hence Set 1 is placed apart in our definition of the self (Set 1 + Sets 2 & 3 + C + B_p = the self).

on all the self-comprising components (and we may speculate that such would in probability hold even if the robotic body were built to look like my pre-transfer body, though to an admittedly lesser degree – the same areas of abilities, presence, expectations, et cetera would be involved).

What would it be like for me to hug my wife and daughters as a cyborg? How would they react when I did? How would their reactions then in turn influence me? How would the absence of the warmth of human bodily contact affect me? How would I mentally handle all of these differences? Although my memories and parts of my self-constituting elements might carry over, my cyborg comprehension of the being I now am (and how I emotionally react to that comprehension (self-comparative and self-conscious emotions)), my new place in the social milieu, and what I seek to do from that altered social position, would be so utterly foreign as to only very tenuously be linked to ‘my’ former incarnation. This conclusion appears to be valid, or at least a probable outcome, if we think back to the divisions we drew earlier between self, personal identity, and whole persons. Personal identity was argued to include the self plus contingent facts about the body, other contingent facts, and social feedback; all of those aspects would be very different for me as a cyborg. To that we added embeddedness to arrive at the whole person; assuming that the operation was not done with many years of cryogenic interval (or the like), the embedded details (historic setting, socioeconomic setting, geographic, climatic, et cetera) might also remain largely the same, but both the self and personal identity particulars would be sufficiently dissimilar as to carry over in significant ways to the whole person of which they form constitutive parts. Thus, we would have a self (and personal identity, whole person) that existed both before and after the brain transfer, but the two could not be said to be the same; the self could arguably be transferred but it could not survive the transfer for any amount of time longer than those first few seconds after being post-procedurally

awoken as too many changes would too soon be demanded. (The reader will recall that we drew a similar conclusion when considering Parfit's teleportation cases in Chapter 1.) What, though, might be the results of a less revolutionary change? What might happen if the brain were transferred into another human body?

As remarked, a body need not be biological in order to give rise to a self on our self-view as long as its (biological) brain is functioning normally, but the type of self a robot body might yield would be quite different, partly due to internally held considerations and conclusions and partly due to how others may be with the being in question. In our new scenario, however, we have done away with the mechanistic and now put our brain into an actual human body. This would mean taking it out of its vat and surgically inserting it into a brainless human form, but if the technology and techniques exist to do all that we have described so far we may as well imagine that this latest version of our musings is feasible too. Since our concerns regarding the self in the previous case rested on the inevitable adjustments made to self-related emotions and reflections that resulted from both fresh internal judgments and distinct social input regarding the new body – with the social side potentially carrying the heavier burden –, we recognize that those same issues would be in play here.

Other people would conceivably react much more warmly to an altered human body than they might to a brain connected to a robot, but there is little difference in kind between this thought experiment and the last one. The revision in treatment this embodied brain-into-person would receive could vary in any number of ways, but it would vary. If, for example, the person suddenly became very beautiful (perhaps the brainless body was genetically engineered and then grown in a lab to perfect physical symmetry and proportions), or very unbecoming (the body came from a terrible accident in which the

limbs, torso, and face were mangled but it was still determined to be a good enough host for our wandering brain), then how others relate to the just-produced brain-body combination could be considerably modified from how they related prior to the procedure, and this in turn would inform the formation of the self as it did for the cyborg. Similarly, the internal reflective issues regarding commitments, traits, aspirations, et cetera (Set 1), and those regarding abilities, appearance, expectations, et cetera, would also require adjustments to greater or lesser degrees as circumstances warranted. Depending on the body involved, the reset that the self underwent may not be as severe as it was in the case of the robotic body (or it may be equally severe, but I see little reason to think that it would be more severe), but it would nevertheless occur, and practically speaking the brain would again redevelop the self from within its transformed circumstances based on the same kinds of interests as in the previous thought experiment. Once more, the self could only be said to be transferable preliminarily; too many adjustments would be immediately required for it to be judged to have survived.

On that note however we should keep in mind that by this account every one of our selves goes through any number of alterations during a lifetime, and that the more minor everyday bodily changes we experience can also affect some of the same features that we have been considering in these more radical cases. As I become greyer and balder my Set 1 commitments, ideals, and possibly even aspirations may largely remain, but my body image modifies and with that my Sets 2 and 3 emotions shift somewhat (primarily regarding the traits aspect of Set 1, but not only that). Reflecting on such a mundane facet shows too how closely linked the self and personal identity are, and how both can feed into one another: while the self is foundational to personal identity the latter's contingent facts and social feedback also loop into self formation.

In summary, the self so far has been found to be real in the way that consciousness is real, as an emergent property of the natural functioning of the brain and fuller body (we will look further into what kind of reality may be claimed for the self in Chapter 6). The self cannot exist without a body of some kind that allows for perception of and action in the world one inhabits, thus leading to thought and that identifiable ‘locus of thought, perception and action’ that the self is (as Cassam put it¹⁶⁰), and moreover is composed of – by the self concept being argued for – Kristjánsson’s three emotional sets adjusted as described along with consciousness and bodily presence. Although the self here is partially comprised of three sets of self-related emotions that stem from and are generated by the individual living in the world and being subject to various stimuli and experiences, it develops in such a way that it becomes something other than the body out of which it manifests. Yet, although it is a something else (in the sense of being a product of natural functioning (emergent property)), the self on this view cannot be transferred: it is bound to the body from which it arises and whose physicality, experiences, reflections, and interactions inform how that self exists in an ongoing process that lasts throughout life.

2.4 A word on consciousness and category mistakes

We have compared the development of the self with how the sense of a unified consciousness (awareness) is believed to form to help explain how the self may be better thought of as an emergent property, and so with the aim of avoiding any unnecessary confusion regarding this aspect of our self-view I think that a brief look at how we tend to speak and write of consciousness is in order, partially to demonstrate the fragility and difficulties involved in its communication, and partially to highlight how the claims we make of it in relation to the self must be done with precision and caution. To explore what is real in a nonphysical way toys with everyday notions and received concepts, the mind

¹⁶⁰ Cassam, *op. cit.*, p. 154.

stretches and does not wish to, and the temptation to take shortcuts, to fog over what should be clear, is always present.

To begin, a recap. Although the picture of consciousness that was provided by Gazzaniga and sketched in Chapter 1 outlines the current materialist consensus view,¹⁶¹ it is not without its detractors; but before glancing at some of the criticisms that have been made of it, a concise synopsis of that present standard account as put forward by cognitive scientists may prove beneficial. The brain, it is put, does not have a central headquarters, a director, an overseer, or the like. Rather the compositional structure is of a multitude of differing specialist systems that each handle specific and small portions of the overall work load, and which are linked together in the neural network and constantly in communication with the other areas. The sense of a unified consciousness (mind) emerges out of all this, but it is more accurate to describe human brains as having a ‘constellation’ of consciousnesses, or consciousness systems, whose various products are integrated and interpreted by a different kind of specialist, a different brain-tool: an interpreter module. As noted earlier, this hypothesized module too is not the ‘head’ or ‘boss’ of the grouping as its role is descriptive rather than generative, the procedures involved in its use are an effective way for us to make sense of our lived experiences.¹⁶²

Now to some critiques. Thomas Nagel, amongst others, has called this explanation into doubt as it cannot explain all that we see and know in the world (e.g. mind, value, purpose, et cetera). This skeptical position argues that since creatures with mind are a natural part of the world what is needed is some kind of a replacement for the purely materialist

¹⁶¹ That is, the general or rough consensus, at least; the whole field of the brain sciences is advancing so rapidly that disagreements of interpretation regarding new data are not always settled before additional discoveries are made.

¹⁶² Gazzaniga, *op. cit.*

understanding; to wit, a biological evolutionary theory that can account for mind.¹⁶³ Nagel thinks that we are missing a large ‘something’; as he puts it, ‘Nature is such as to give rise to conscious beings with minds; and it is such as to be comprehensible to such beings.’¹⁶⁴ To Nagel, the notion that a mental state can be reduced to a corresponding physical event in the brain cannot answer what property a mental event would have that would be distinct from the physical properties of the associated event, leading advocates of this view to retreat into behaviorism. This in turn is faulted as being verificationist (i.e. all that needs to be said about a mental state is what can be attributed to it via an observer), leaving out the first person point of view – what sugar tastes like *to me*, what anger feels like *to me*, et cetera.¹⁶⁵ This is important, Nagel says, because the reductionist model is based on concepts like ‘water = H₂O’ which can be measured and thereby fully realized regardless of an experiencer, a subject. However, experienced things like taste or feel do not seem to match so evenly with a brain state, they seem instead to be ‘something *produced* rather than constituted by the brain state. So it cannot be identical to the brain state in the way that water is identical to H₂O.’¹⁶⁶ Emotions may provide a clear example here: we can claim that a chemical like serotonin makes us feel happy (or, more accurately, contributes to that feeling), but that does not tell us what happiness is like nor how my happy occurrence differs from yours when each is felt from the inside, as it were. So far, I find this to be a compelling argument – worth much thought – and we will return to considerations such as these in great depth in Chapter 5.

Returning to our recap and its critiques, the neat reduction of a lived effect to a purely physical cause may leave out the related first person perspective (at least descriptively), but

¹⁶³ Thomas Nagel, *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature Is Almost Certainly False* (New York: Oxford University Press, 2012).

¹⁶⁴ *ibid.*, p. 17.

¹⁶⁵ *ibid.*

¹⁶⁶ *ibid.*, p. 41.

that is not all that could arguably be found lacking. Readers will probably have further doubts. What, for instance, does it actually mean to call consciousness an ‘emergent property’? Definitionally it indicates complexity born of the interaction of simpler properties, yet the claim made regarding consciousness on the standard model is that many individual and locally-focused consciousness systems together generate a sense of a unified consciousness likely with – but possibly without – the aid of a specialist ‘interpreter’ module. Fine and good we might think; this is then clearly a case of emergence. Yet what are the simpler properties that are working together? Local consciousnesses? What exactly are they? Furthermore, how do such differ from the overarching and ‘unified’ consciousness? Much seems to be taken for granted on this account as at present we have little that can be pointed to physically as being this or that. Based on evidence from techniques such as fMRI scans experts can state that this area of the brain is associated with X or Y observable behavior or process, but a brain is of course unlike a clock where the cogs fit together distinctly and are easily separable. Moreover, simply applying the label ‘emergent property’ and then moving on is a bit like, in this case at least, making use of a ‘get out of jail free’ card: it reads a little too much for comfort along the lines of: ‘We know that consciousness is there but cannot really explain how (or why) and so we will just say it is emergent.’ As Nagel points out, that is not really enough. Yet that judgment too might be premature, for as Gazzaniga would counter, no one is simply moving on either, the problem is receiving much attention. (Our study will also later examine the physical basis side to consciousness, and reasons for thinking a physical basis existent are there and describable in some detail.) Perhaps therefore the most we can do at this point in our present study is to admit that we do not yet have the full story, and Nagel’s point, if I read him correctly, appears to be that in looking for the whole story one ought to widen the search beyond the purely material. To what is an open question.

Nevertheless, the model as it stands does remain useful in our investigation into the self for a number of reasons. Firstly, while it may not (or it may, we refrain from taking a stance until Chapter 5) contain all the parts of the consciousness puzzle it does contain many more than have previously been available and we therefore ought to take what we can from it as we continue. Secondly, in conjunction with our earlier argument, part of what we can take from the standard conception is the idea that since consciousness is better understood as a network of local systems (a ‘constellation’), consciousness is not something one ever loses as long as a minimum of one of those systems in the network is functioning. Understanding consciousness in this way added an important aspect to our definition of the self. Thirdly, ‘emergent property’ is not an empty phrase and does carry an explanatory load. Admittedly, employing it in regards to the self is in some ways even less instructive than it is for consciousness (with consciousness we need only look at the brain whereas with the self we apparently must look not only to the brain but also to the whole body, and then not only there but also to how that whole body exists in the world in which it finds itself), but it does help us picture how a self is a part of a whole person without being physically definable, and how a self can be substantive without being locatable. Finally, from Nagel’s objections to the standard model we are reminded of the centrality of personal perspective. That experience, one’s personal experience, the first person point of view, is a crucial part of what goes into the self should not be disputed and will be discussed at length in the next chapter. However, for the time being it is sufficient to call attention to the fact that consciousness should not be automatically equated with mind (nor awareness), and that mind in turn should not be automatically equated with the self. The situation is (unfortunately?) much more complex than that.

Such mistaken identifications have come to be referred to as ‘category mistakes’,

following Gilbert Ryle's exposition in *The Concept of Mind*.¹⁶⁷ Ryle writes that to confuse brain and mind is to look for mechanistic workings of the mind; this may apply to the brain (we assume that it may apply, Ryle does not explicitly state so), but uses the wrong approach and the wrong vocabulary for mind.¹⁶⁸ This is a relevant point to bear with us as although physical processes can reveal much about how the brain functions, mind – whatever that is – very likely plays by different rules, or at least acquires something extra in the way the parts playing by those rules interact, hence its description as an emergent property. The self too, if our argument carries at least to the extent that it can be thought of as an emergent property, should be approached with linguistic (and therefore perform conceptual) caution as ordinary language can so readily lead us astray. Ryle points out that, 'the sentence "Cremate me after I am gone" says nothing self-annihilating, since the "me" and the "I" are being used in different senses.'¹⁶⁹ This is a striking way of pointing out how it is all too easy to use words like 'the self', 'myself', 'me', et cetera in ways that blur the focus of what we are trying to explore. While I have sought to maintain a rigorous approach to the use of such terms I am mindful that failure or slippage on my part is likely, and that – as just shown in our consideration of the term 'emergent property' – a definitive answer to the problem of the self seems a long way off yet. I would therefore ask the reader for continued patience in the remainder of this chapter, and to bear in mind the core definition of the self that we offered, as well as its two-way relationship to personal identity and whole person.

2.5 Certain uncertainty, randomness, and limited choice

This will be the final explanatory section of the core of our self concept before proceeding on in the next chapter to consider some possible objections to it and replies to those

¹⁶⁷ Gilbert Ryle, *The Concept of Mind*, intro. by Daniel C. Dennett (London: Penguin Books, 1949, 2000).

¹⁶⁸ *ibid.*

¹⁶⁹ *ibid.*, p. 180.

objections, along with a look at some of the wider considerations involved from a phenomenological perspective. Our objective in this section is to properly place the self in its world, to situate it in its embeddedness and to explore what such might mean for our theory of the self.

Richard Rorty has pointed out that ‘the central flaw in much traditional moral philosophy has been the myth of the self as nonrelational, as capable of existing independently of any concern for others’.¹⁷⁰ We have tried to avoid this error by situating the self socially in the way that others affect how we come to understand ourselves, the social feedback loop that influences self-comparative and self-conscious emotions (Sets 2 and 3), which in turn go into how self-constituting elements (Set 1) are formed and reformed over time. However, this is not enough to account for the whole of the self and its wider placement, and as I am sure Rorty would agree, any given self is also highly dependent on the historical, socioeconomic, geographical, epochal, and other elements that accounts like Ravven’s highlight and which were laid out in Chapter 1. The spaces in which we find ourselves, in which our selves exist, are marked by many outside pressures and unacknowledged internal inclinations; our selves are not only interdependent with each other but also with the environments in which we are embedded. Our account thus far has yet to situate the self in this contextual manner and examine both how we see the world and how we fail to really see the world; that will therefore be the focus of what follows.

The general picture we tend to have of ourselves as being rational agents calculatingly making choices based on all of the available information in an open and free world was challenged by the evidence from psychological studies like those described in Haidt’s

¹⁷⁰ Richard Rorty, ‘Ethics Without Principles’, in *Philosophy and Social Hope* (London: Penguin Books, 1999), pp. 72-90 (p. 77).

work,¹⁷¹ data that resulted in the two-tiered behavioral structure (intuitive first and foremost, and then (sometimes additionally) rational) outlined in Section 2.2 above. That we have, for a number of interesting but distracting (from our study's purpose) historical reasons, come in the post Enlightenment period to be inclined to apprehend ourselves as logical decision-makers and as unhindered in the options available from which we then purportedly choose has consequences for the (often unexamined) outlooks adopted towards our lives and our persons. If in fact we do not make choices in the way that we think we do, and if in fact we are not as 'free' (in the sense of being unencumbered when deciding) as we think we are, then how ought we to adjust our conceptual and attitudinal stances and what might be the consequences of such an adjustment? In all honesty I am not confident that I can give a fully clear picture of the view I wish to argue for in this section as in many ways it cuts deeply against the grain of received thinking (and is therefore also counterintuitive regarding many of the taught and entrenched automatic judgments we likely hold), but my utmost efforts towards that end will be made. To start we will need to take another look at the psychological processes involved when we do make choices, and we will moreover have to re-examine the world we find ourselves situated in, analyzing a side of life that is seldom given attention in the everyday bustle of activities. What we find will be disturbing.

By way of background to the arguments that will be made in the following, to the psychological and cognitive issues that have already been discussed we can add Daniel Kahneman's research specifically into decision-making, studies that resulted in the proposal that our cognitive processes are composed of two separate, but linked, systems: System 1 which is intuitive, fast, automatic, and gives impressions of percepts and thought objects, and System 2 which is rational, slow, laborious, and whose analyses are always

¹⁷¹ Haidt 2001, 2012; *op cit.*

done purposely and explicitly.¹⁷² Kahneman notes that most of our behavior is intuitive, skilled, and successful, and that even when not directly intuitive it is ‘anchored in intuitive impressions and intentions’.¹⁷³ As the reader will have realized, this lines up neatly with Haidt’s behavioral model, and in fact Kahneman cites Haidt’s work to this effect, along with Gary Klein’s.¹⁷⁴ Thus, via this additional body of work we again have further empirically-grounded evidence that propounds that the majority of the time behavior is automatically generated and therefore not explicitly planned, that expressed outputs are the end product of a built-in system (System 1) that reacts to stimuli and perceptions in the external world in a manner that (more often than not) is advantageous, or at least neutral, for the agent involved. The biological basis and the ‘reasoning’ (from an evolutionary perspective) behind such a system should be clear; in addition to the increased organism preservative features of rapid decisions and responses the far greater efficiency found here also has important energy-saving advantages, amongst many other benefits. The non-automatic System 2 that we, and to varying degrees other advanced nonhuman animals, have can intervene but it must be ‘turned on’, that is, it must be willfully accessed and employed, and even when so engaged its workings will retain as influences and prejudices the results of the preceding (and unavoidable, not ‘turn off-able’) System 1 processing. It seems that we can never get away from the intuitive judgments that have been highlighted, nor the emotional input that supplements them, the latter in both its similarly inbuilt and learned varieties.¹⁷⁵

That we therefore act, and even choose how to act (or, perhaps more accurately, think that

¹⁷² Daniel Kahneman, ‘A Perspective on Judgment and Choice: Mapping Bounded Rationality’, *American Psychologist* 58:9 (2003), 697-720; see also Kahneman’s *Thinking, Fast and Slow* (New York: Farrar, Straus and Giroux, 2011).

¹⁷³ Kahneman 2003, *ibid.*, p. 717.

¹⁷⁴ Kahneman refers to Haidt 2001, *op. cit.* and Gary A. Klein, *Sources of Power: How People Make Decisions* (Cambridge, MA: MIT Press, 1998).

¹⁷⁵ These categories of emotion, and the distinctions between intuitions and emotions, were discussed in Section 2.E above.

we choose), based largely on automatic and preconscious conclusions regarding the perceived and experienced world would not perhaps warrant much concern if we were adequate judges of the situations and circumstances we found ourselves in, but we tend to be quite poor at recognizing the important role that chance and probability plays in our lives, and this affects our ascertainments in many and far-reaching ways. In the following we will therefore attempt to look at the world that scientists and mathematicians maintain we actually do inhabit, its forces and its structural frames, and not the world that we – that you and I – might prefer to think we inhabit.

Those of us who are neither trained statisticians nor deeply engrossed in probabilistic reasoning usually (erroneously, as it turns out) assume that the results of our actions and the situations we encounter follow naturally from the choices that we have made, and this is due in part, evidence indicates, to how we are biologically wired. Amos Tversky (Z”L) and Kahneman’s groundbreaking 1974 study on heuristics and biases revealed a number of unsettling conclusions about how we human beings consider the events we witness happening around us.¹⁷⁶ For one, the authors report that we tend to be overconfident and overly positive in our determinations of the world we live in. When ascertaining an event or circumstance’s likelihood we are prone to consider its probability to be high when we think it is representative of or similar to something else and low when not, and whatever familiarities we happen to have – with the arbitrary sources of (at least some) such certainly coming into play here – quite naturally affects these judgments and thereby further distorts thinking. Yet determinations of these types (overconfident, overly positive, high/low likelihoods, et cetera) ignore important factors such as sample size and prior probability, amongst others, and many of us in our daily lives are no doubt ill or even

¹⁷⁶ Amos Tversky and Daniel Kahneman, ‘Judgment under Uncertainty: Heuristics and Biases’, *Science* 185:4157 (1974), 1124-1131. Tversky passed away in 1996.

uninformed about the importance of these elements and the implications they carry. This is consequential, since mathematically speaking representativeness and/or similarity are unrelated to probability, and moreover, as political opinion polls often reveal, small sample sizes frequently yield misleading and unreflective results.

Furthermore, adding to our (mis)interpretative worries, we tend to misjudge series, expecting that ‘a sequence of events generated by a random process will represent the essential characteristics of that process, even when the sequence is short’, leading to error as ‘a locally representative sequence...contains too many alternations and too few runs’.¹⁷⁷

In essence, the researchers are explaining that we see a tiny part of the picture and think that we understand the whole thing, failing to appreciate the randomness that inevitably is involved in any complex, real world situation, simply scaling (or inflating) out what we can readily or easily observe – and we cannot readily or easily observe whole pictures. This problematic penchant is comparable in its effects to the problem of small sample sizes. In his book describing this same issue of wrongly approaching series and similar tendencies that even the very educated amongst us share, Leonard Mlodinow asks if a million consecutive zeroes (in a binary system), or the success of Wall Street prophets, could be generated randomly – the answer to both, incredibly (for the former at least; the latter may not surprise the more economically informed), being yes.¹⁷⁸ Please take a moment to think about that. In binary numbering, such as that used for coding computers, only either a zero or a one are ever put to use; say that I wish to generate one of the two digits randomly, which number will be produced? Quite obviously the individual odds of either are fifty percent each time taken singly. Yet one million of the same in a long string

¹⁷⁷ *ibid.*, p. 1125.

¹⁷⁸ Mlodinow’s book is both a very readable and a very informative account of the importance of randomness explained using approachable mathematics: Leonard Mlodinow, *The Drunkard’s Walk: How Randomness Rules Our Lives* (New York: Vintage Books, 2009); of particular relevance to the present is Chapter 9: Illusions of Patterns and Patterns of Illusion, pp. 169-191.

can be generated randomly. Now try to visualize that, here is a small sample to help:

00

00

We would need 10,000 more of those strings, for the above is only one hundred zeroes in a row; and while I of course purposely typed all of those out for the sake of demonstration, in a random process there would be an even chance for each and every one of those zeroes to have instead been a one. Now again multiply the string by 10,000. It is incredible. This type of thinking is so counterintuitive that it feels difficult to even wrap the mind around. That is a necessary point for us to stress: This type of thinking is *counterintuitive*. We are not naturally prepared to accurately judge in these terms – to do so requires years of specialist training – and since so many of our judgments are automatically and preconsciously generated the position we consider ourselves to have within the world we inhabit appears to contain a deep disconnect between how we view our environments as functioning and how they – in global terms – actually do function. Perhaps our biology has blessed us with skills that are ‘good enough’ for the level and terms on which we daily struggle, but our universe is a thoroughly interconnected one and our eyes simply do not perceive that. To put it prosaically, we are not prepared to see the forest for the trees.

Moreover, as if the preceding were not enough, we also consider that an event will be more frequent or probable when we can more easily recall an example of its class or an analogous occurrence, and we overestimate the likelihood of conjunctive events (such as a plan or project in a series of steps) while underestimating the risk involved in disjunctive events, where only one aspect slightly off would have deep consequences (such as a

complex system's failure caused by one faulty part).¹⁷⁹ All of the above taken together indicates that we are apt to be overly optimistic that we will be able to actually accomplish what it is that we set out to do, operating – in our decision-making – from the narrower view of gains and losses, yielding a perspective that is much further from a risk neutral position than a more balanced and inclusive point of view would provide.¹⁸⁰

The results of this research are essential to our attempt to contextually place the self because they tell us that we generally look at the world, and our individual positions within it, through roseate lenses, and this viewpoint quite naturally has far-reaching influences on the self-related emotions that we have been examining. For example, as a general rule we typically fail to appreciate the extent to which uncertainty marks our lives and the dominant role that randomness plays both in what happens to us and in what results from the things that we set out to do.¹⁸¹ When something good occurs we laud our abilities, and when we fail we blame our ineptitude, in each case generating self-conscious and/or self-comparative emotions (Sets 3 and 2, respectively), depending on whether or not we are measuring ourselves against what we 'should have' been able to do or 'should not' have caused. Additionally, we frequently take no heed of the deeply rooted affective influences merging with our intuitive interpretations and reactions, not even recognizing that such exist as we internally (and, if called upon, externally) explain our own behavior by affixing *ex post facto* reasons to what we notice we have just done (a process outlined in Chapter 1). Again, these are actions that were generated preconsciously and not those following a

¹⁷⁹ Tversky and Kahneman, *op. cit.*

¹⁸⁰ Kahneman, *op. cit.*

¹⁸¹ Perhaps there are some people who do take this into account and have multiple alternate plans for each scenario they can imagine occurring – such as one might have in the advanced stages of a game of chess – but even if such people exist, and the demands of the everyday being what they are for the average person I think such ultra-planners must be very rare, it is impossible for anyone without omniscience to foresee *all* of what might happen and then to create the appropriate alternative stratagems.

rationally considered decision.¹⁸² We tell ourselves that ‘I did A because of B rationale’, when in fact A was caused by a chain of automatic processes that needed no input from reasoning faculties.

Although the foregoing has admittedly contained many generalities, I ask the reader to remember when considering the overall view here summarized that such statements are the conclusions of numerous empirical studies and so we ought – at least arguably – to take them as being just that: currently accurate accounts of contemporaneously accepted scientific studies which are potentially subject to change (or to being further reinforced). Descartes described nonhuman animals as being like machines, unfeeling and unthinking near-automatons that responded in set ways to perceptions and stimuli;¹⁸³ it can be very unpleasant to learn that this description applies just as well (but also just as poorly) to ourselves. Of course neither we nor our nonhuman compatriots are unfeeling and unthinking, but there are few differences in the manner in which we customarily respond to our worlds. Note though that this summary has mostly ignored the deeper inner life; what we have been discussing are default decision-making processes and orientations of perspective. This has largely disregarded too the System 2 suite of rational tools that Kahneman, Haidt, and others describe. Let us try to further apply what we have learned.

The psychological and other research here related regarding the broad trends of how we operate in our environments is focused on the surface and does paint with extensive strokes, but nevertheless what it reveals is instructive for our concerns with the self. We have read that we are such as to be unaware of, or to ignore, the internal and external factors in our lives that directly affect how we understand where and who we are and how we relate to

¹⁸² Kahneman, *op. cit.*, Haidt 2001, 2012, *op. cit.*; see also Greene *op. cit.*; Chapter 1, Section 1.1.C (‘Four accounts of the self: The soft realist position’) contains a brief summary of the general process as well.

¹⁸³ Descartes, *op. cit.*

the others around us.¹⁸⁴ We are likely to think that we are freer to act and to choose than is the case because we are not usually informed of the extent to which these psychological (and biological) characteristics limit what we might consider our options to be, nor do we comprehend just how much of the ways that we do act and do choose have been preconsciously determined. This is important because when we seek to employ our rational abilities we will not be doing so in a vacuum: such will be informed by any number of preconscious inputs and affective influences. Pure objectivity, truly *rational* analysis, might be close to impossible, though we can perhaps hope that a philosophical methodology may be able to take us close to that goal. (Should we want to get there – another worthy philosophical query.) Again, how all this relates to a self born of our three emotional sets, consciousness, and bodily presence should be clear enough: one’s internal positioning will necessarily affect all areas of the self-constituting, self-comparative, and self-conscious emotions and from there to personal identity and whole person.

Additionally, although we may tell ourselves – or at least be tempted to tell ourselves – that we can go out and do anything, *be* anything, that our hearts desire (a common trope of US education), what our hearts actually will desire is embedded in a network of cultural prejudices and concerns, inherited values, genetic dispositions, sociohistoric worldviews, et cetera, et cetera. If the facts of this matter are not faced then in the forming and shifting of one’s Set 1 self-constituting ‘emotions’ one may well be condemning oneself to much frustration. We are not fully free to do absolutely anything at any time because what we will be able to do, and more to the point, what we will consider ourselves able to do, will be limited by factors over which we have no control. This is a nuanced point and warrants some further reflection. The aspects of our natures and our situations that impose

¹⁸⁴ This includes, by the way, aspects of how we fared in the genetic lottery; we may realize that others are prone to treat us well if we are very good-looking, but do we also realize that our disposition to mistrust and to doubt that tendency has roots in our upbringing and perhaps also partially in our DNA structure?

constraints typically occur beneath awareness, are ongoing, and have generated their output before we are able (using System 2) to process the input they might provide. That is, such give us ready-made answers and, at least as far as behavior is concerned, *fait accompli*. This is Kahneman's System 1 (and Haidt et al.'s intuitive judgment) in action, and System 1 is always in action. The blinders nature has given us are composed of these automatic evaluations regarding stimuli and interpretations regarding perceptions, such as the physical and emotional responses described above when confronted by a snake. Our blinders do not end there though, as our embedded determinations also play their part (cultural, epochal, geographical, socioeconomic, et cetera). This might seem obvious to many readers, but the ramifications for the self are important and so let us extrapolate a bit more on the matter.

If, as has been argued, the self is a real and ongoing creation of a whole person moving and being in the world, then the parameters that determine the contours of that self will be crucially important, and what those boundary lines appear to indicate is that we normally cannot see the full options that we really have before us, and in that failure we may limit and shrink ourselves. We seem incapable of understanding our own freedom and thereby stand to lose a great deal of it. The contemporary economic discourse is a good example here: Does anyone in the developed world of 2018, even those deeply dissatisfied with the current state of affairs, really think there are worthy alternative economic models to market-based capitalism? Some few do, of course, but that they are so far from the mainstream that they are immediately dismissed (or, more potently, ignored outright) tells us all we need to know about this facet of our *Zeitgeist*. Yet it is just that: simply one part of a current trend that will eventually give way to another, although one would not really know that by reading or listening to the pundits, who either lack the imagination for what else could be or have arrived at the conclusion that nothing else *could* possibly be. Either

way the case in point is demonstrated – potentialities are whittled down before they can reach us. In a self that is ever-built potentialities are everything.

Yet the preconscious, which does so much of this paring, is an amazingly efficient information processor, far more so than our conscious minds.¹⁸⁵ Should one therefore make all one's decisions based on one's gut? That is a moot question precisely because decisions will often already be made before one even (employing awareness) consults said gut. If we do find ourselves operating System 2 and rationally chewing things over, the morsels of information we will be nibbling away at will include influences that are both intuitive and affective – seasonings, of a sort. That problematic bit of 'asparagus' we have been working on is not as raw as we might (rationally) think it is. Nevertheless there our rationality is, offering the chance to override what has been delivered via automatic processes, offering the power to choose: for better or for worse. Again, in a self that is always in motion, always growing, the consequences are manifest.

The reader will remind me that I have not denied choice, only asserted that what choices we consider ourselves to have will be limited. The influencing factors on our thought are unavoidable – inescapable – but we are not automatons, and if we determine to do so we can recognize and even adjust the intuitive judgments that are provided through purposive reflection on how we tend to react and respond (concerning both thoughts and actions), awareness-based repetition of desired outlooks or behavioral patterns, and/or outside influences from the social sphere as per Haidt's model.¹⁸⁶ One's rational powers are naturally very helpful in this. While intuitive judgments, affective influences, and circumstantial pressures (regarding the shallow sense of situation and also the deeper sense

¹⁸⁵ Dijksterhuis, *op. cit.*; for the discussion on this see again Section 2.2.E above ('Intuitions and emotions').

¹⁸⁶ Haidt 2001, 2012, *op. cit.*

of embeddedness) may be pushing in a certain direction, we still have the ability to choose within the framework that we confront; or, indeed, to choose to work on changing the framework itself. To this glimmer of hope, however, we must add the processes embodied through the forces of uncertainty and randomness, elements which we cannot even dream about mastering (realistically anyway; we could *dream*), and which are not usually taken into consideration when we make plans and plot out what to do when in order to get where we think we want to be. In sum, we typically fail to recognize both how restricted what we think our choices are, and how narrowly defined they actually are through the degree to which the results depend on chance occurrences and outcomes that are only probable to varying magnitudes. Our selves are situated in a many layered complex of mutually dependent and interlocking influencing factors, which points to a limiting of potential actions and options, and this the more so when we consider that in most cases our preconscious minds will be the deciding element, each time with internal influences of current affective aspects and the aforementioned societal concerns. Moving through all this with the first person perspective that is innate to us it is plain to see how we can miss so much and mistake our current concerns and perceived options for being all that there really is. On these shaky grounds our selves are formed and we continuously re-form them. I do not know if there is a good or even adequate response to this; it might just be a part of the human condition. Still, I believe that we can and should employ our rational abilities to think about how we want to think about all this – to the extent that we are able.

2.6 Taking less (but taking)

It can be hard, at least it is for me, to take all of the above on probability, randomness, and limited choice in at a deeper than superficial level because we are so accustomed to our inner lives and inner voice – assigning reasons, rationales, weighing pros and cons. Yet as the accumulated evidence and our analyses have indicated, that voice is the overdubbing

on a soundtrack that has already been played. When we examine our everyday movements we are able to recognize just how little reasoning (System 2) is involved in the determinations, behaviors, and decisions that manifest themselves. Haidt even mentions that trained philosophers are perhaps the only group of people that really regularly put their rational faculties to good use¹⁸⁷ (one more reason we make terrible dinner guests). Still, just how helpful all the above is for philosophers, or anyone, in purporting themselves is surely a wide open query. The burden now presents itself of how one ought to adjust one's life in light of the above – knowing this, what are we to do? It seems like the type of knowledge that enchains rather than liberates: we are at the mercy of unreachable preconscious processes and the unpredictable random consequences of a natural system far too intricate to fully comprehend.

This certainly appears so, but the situation does allow for some benefits to the individual when approached from that same first person point of view. There is, initially, gratefulness for the disasters that have not yet hit us (as Mlodinow puts it¹⁸⁸), and stemming from that feeling it may be possible to cultivate within oneself gratitude for simply being alive, even if we are not exactly the type of creatures we may prefer ourselves to be. What I think may be more perspectively helpful however, and more deeply rewarding from a whole person (embedded) standpoint, would be something akin to the postmodern reaction to modernism, that is, to a certain embrace of ambiguity,¹⁸⁹ to accept that very much of life really is like throwing the dice in a dimly lit casino. Although we can have no idea how things will turn out for us as we strive towards our goals (given randomness and compound probabilities), and although many of the things we decide and do will not turn on what might be labeled 'our' choices (i.e. those arrived at via aware and internal deliberation), still we can perhaps

¹⁸⁷ Haidt 2001, *op. cit.*

¹⁸⁸ Mlodinow, *op. cit.*

¹⁸⁹ John D. Caputo, *Truth* (London: Penguin Books, 2013).

take meaning from the efforts themselves and accept our preconsciously made decisions as being expressions of the selves we are, that is, as being products of who we are as a functioning human being in our time and place. We can look at the products of our intuitive judgments as potential lessons about the nature of our selves (Set 3: self-conscious emotions), as opportunities for increased self-knowledge. Such might then assist in the construction and/or maintenance of one's current Set 1 self-constituting elements. If the self is real in the way the foregoing has argued it to be then internally directed reflection of this sort seems entirely warranted, and may even be – dare I say – a rational attitude to take. Moreover, there is also the sense in which merely gaining this type of knowledge does increase one's control over how one operates in the world: by forcing attention onto the automatic one gives oneself the opportunity to interfere/intervene, to apply System 2 to try and shift what System 1 yields by adjusting intuitive judgments and affective reactions. Yes, we cannot really realize much about our world and what will happen to us in it, but we can learn about how we, as one person, are in that world and how we have been shaped by the forces of that world.

In gaining access to this we can thereby start to think about, and too start to create, the type of person we would rather be given what we have come to understand. This self-transformation would begin with the converting or erecting of one's self-constitutive elements (Set 1) while monitoring the project with the help of one's self-comparative and self-conscious emotions (Sets 2 and 3). What I am suggesting is a program of calculated self-development – that is, literally development of the self – that, from there, has all the related extensions into personal identity and whole person issues. Much of how an individual involved in such a project would choose and act would still be resultant from the automatic processes we have examined, but the included thoughts and behaviors would at least be monitored. If one wishes to be more, or to be other, than one is such may be the

place to start. Once acquainted with how defined the unexamined self is one may be better positioned to employ some imagination and become differently, to see past the circumstantial (in the deep sense) barriers and seek alternate ways of being. The key to unlocking all this might very well be simple familiarity with one's default modes. We are all seemingly unavoidably limited, but the research we have surveyed and the conceptual applications we have arrived at indicate that we can at least have some say as to the extent.

This discussion of self-becoming opens up another avenue of interest: since the self on this understanding is a part of personal identity and therefore also of the whole person (and due to its central role a critical part), what would an alternate self-view mean for that whole person in the world? If the foundation were re-laid, how would the shape of the structure differ? These are questions worth keeping in mind as we continue our study on the self, beginning in the following chapter with some possible objections and replies to the theory of self outlined here, focusing especially on phenomenological issues.

Chapter 3: Phenomenological Approaches to the Self: Objections, replies, and objections

3.1 Review and preview

In the preceding chapters we have argued that the self, personal identity, and whole person considerations should be viewed as standing in a layered relationship, with the self as foundational. On this account the self is composed of one's central commitments, traits, ideals, goals, characteristics, and similar elements, some of which have been determined by factors beyond anyone's reach such as genetics and upbringing, but many of which are nevertheless potentially alterable through the purposive training of altered and/or new personal traits, views, evaluations, and the like (Set 1); to this core we added emotional and intuitive reactions (which are also trainable, as Haidt's and others' work have shown) based on what one would like one's fundamental constituents to be and on how one sees them as currently being, run on an internally engaged self-monitoring and self-critiquing feedback loop (Sets 2 and 3); these identity and emotional aspects are further supported by consciousness (which has been proposed to be ongoing and without breaks; symbolized as C) and the presence of a physical body (B_p), and thus for the three set-based components psychological and physical continuity is maintained through constant renewal. These findings were summarized with the following equation: $\text{Set 1} + \text{Sets 2 \& 3} + C + B_p = \text{the self}$. All of these factors have been determined necessary for self formation (at least in the human case, which is all that has been considered and which will remain the focus of the present study). To this self are added bodily contingent and other contingent facts (details over and above mere physical presence), along with feedback from the social realm (separate from internal feedback but importantly influential in ongoing self regeneration), to yield personal identity, and then finally the entirety of the above – that is, the full

‘package’ represented by personal identity – along with the insertion of embedded environmental factors (‘enworldedness’) produces a whole person.

To speak of a human ‘me’ we must combine all of the elements (characteristics, traits, ideals, goals, results of introspection and monitoring, et cetera) that go into that ‘me’ with the effects of social interactions and of existing in one’s world. The resultant (full) ‘me’ is thereby something that is continuously adjusting and shifting, yet that still seems to carry a substantiveness when viewed from within or when acknowledged in/as another: a recognizable ‘me’ (or ‘you’) that is unique and identifying and crucial to the whole human creature one is. This is a self that is ever in the process of creation and transformation, and yet within that movement still remains constituting and whole at each step along the never-ending way. The manner in which this self concept differs from previous theoretical accounts is noteworthy and bears highlighting: whereas other self-views in both academic and religious contexts (and their overlap) generally take the self as either being nothing at all or as being something that exists antecedently and needs to somehow be ‘found’, this concept has the self as arising and as becoming – gradually and unceasingly – out of a being living in a world. It might seem counterintuitive, and indeed definitional folly, but there is a sense here in which the self rests on the whole person just as much as the whole person rests on personal identity and on the self.

In a discussion of G.W.F. Hegel’s notion of the world-historical spirit, John D. Caputo compares truth as becoming to ‘the way the acorn becomes an oak or the infant an adult.’¹⁹⁰ I do not consider the self to be relationally like this, but I bring up Caputo’s point in order to indicate a certain way of approaching the self that might be misleading. I think this comparison, in fact, is illustrative of how we ought *not* to think of being-as-becoming,

¹⁹⁰ Caputo, *op. cit.*, p. 148.

particularly when it comes to a topic such as the self, for it cannot be that an acorn could be said ‘to be’ an oak, nor an infant ‘to be’ an adult. Developmental trajectories do not indicate being-in-time (being-now) and – again, most certainly in the case of the self – there is no endpoint to be arrived at. An acorn is only ever truly an acorn, and as it develops it takes on newness in being, carrying over what was into what is, yet only ever as that *is*, as that being-now. Becoming an oak does not ‘fulfill’ the acorn, it is fully and essentially itself at each stage; in a similar way the self as understood on this account is at the same time a process and an entity.

In the next chapters we will look in more detail at how the self may be that latter (an entity) in a way that is both realist yet non-pinpointable, but prior to that in the current chapter we will focus on the former, on the self’s aspect as a process. To do this we will first consider a very general objection to the self concept as given here and a reply to that objection, and then we will spend the remainder of our time on phenomenological issues and some possible counterarguments and further considerations relating to the proposed self that might arise from those quarters. It is hoped that by doing so we will be able to appreciate the central place that experience and awareness play in the self’s formation and its being. We will initially consider Galen Strawson’s thin self/minimal subject account, followed by Barry Dainton’s phenomenal self concept, before then turning to a review of Edmund Husserl’s phenomenological theory to resituate ourselves methodologically, and thereafter finally, and briefly, consider what Martin Heidegger wrote about the self. I wish to include that last section in the present as some of the metaphysical proposals regarding the self to be expounded in our later chapters will take ideas from Heidegger’s work. It is hoped that although the following could hardly be said to be a worthy examination of phenomenological studies on the self it will be sufficient to at least provide some footing. To commence then a very general objection.

3.2 A general objection and reply

Although some arguments against a soft realist view of the self have already been covered in our consideration of the representative anti-realist accounts, perhaps the most direct objection that could be made at this point would be that it is simply one more theorized account among many in the seemingly endless discussions on the self, with no real reason that would cause us to consider it more accurate than its rivals. In previous chapters we considered psychological, cognitive, and bodily-related issues and concluded that under analysis those elements did appear to point towards a soft realist interpretation, but we can imagine that the same considerations could perhaps be used to argue for other self models. Additionally, the reader may even be questioning the very basis of the entire enterprise: founding a self concept on something as apparently ethereal as human emotion and adding to it the complicating factors of consciousness and bodily presence as necessary parts – it might seem like overkill, madness, or worse yet: not even worth bothering about. What could be said in response to such an accusation?

While our adjusted versions of Kristjánsson's three sets of self-related emotions that compose the soft realist self may be difficult to ascertain empirically rather than experientially, the shifted foundation that was given to our model (establishing an emotionally-grounded reason rather than the other way around) in light of the research on the causal roots of manifested behavior and the means by which decision-making tends to happen (i.e. intuitive reactions primary, rational thinking secondary) does have a sound basis in the experimental sciences, and this might give us pause or at least open the door for further investigations along the lines so far explored. The empirical results in question have also been repeated and promoted by a number of researchers, as the above has outlined, and this too reinforces a certain weightiness to the conclusions we have drawn.

There is moreover a rich theoretical history behind the general view being presented here, even if the details are recent, or, in many instances, newly established. Greene notes that ‘From a neural and evolutionary perspective, our reasoning systems are not independent logic machines. They are outgrowths of more primitive mammalian systems for selecting rewarding behaviors – cognitive prostheses for enterprising mammals. In other words, Hume seems to have gotten it right.’¹⁹¹ (As for what specifically the ‘right’ in question is, Greene states that he is referring to Hume’s comment on reason being a ‘slave of the passions’.) Since our ability to reason, and likewise all that we take to be quintessentially human about that, has at its core an affective network of action promotion, similarly grounding an internally held view of self in the emotions would therefore appear to be far more evidentially valid than many of the alternatives, and unless (until?) a hard realist self were to somehow be discovered these thoughts seem to present us with the proffered view as an avenue at least worth pursuing.

We can still, however, imagine potentially quite different ways that a self could be realist in a soft sense, and even the soft anti-realist accounts we looked at seemed arguably capable of falling into either broad camp regarding realism. One such interesting self-view that skates this line between realist and anti-realist, and that challenges us in our conceptions of the real, was mentioned in Chapter 2: that of the minimal self/subject, also called the thin self or *sesmet*. It suggests that we need not apply any level of realism beyond the experience of the present moment, but that that experience is nevertheless sufficient for a realist self, albeit an extremely short-lived one. This is the first phenomenological self account that we will consider.

¹⁹¹ Greene, *op. cit.*, note to text on p. 137, p. 368.

3.3 Phenomenologically-based accounts and related issues

A. Phenomenology 1: Strawson's thin self/minimal subject

Aaron Henry and Evan Thompson have noted that there is a tendency to lump the phenomenological question about whether or not every conscious experience has an 'owner' (self/ego) and the metaphysical question about the self's reality (or not), 'an answer to the metaphysical question of the self's existence presupposes an answer to the phenomenological question about how experience is structured. In this way, phenomenology is taken to guide metaphysics.'¹⁹² The view, on one hand, that the self or subject is an embodied one, and the view on the other that it is a constitutively bodied one, are both distinct notions yet – not unlike consciousness-mind-self issues – are nevertheless concepts that are easy to blur together.

The metaphysical question regarding the self has usually been answered through the lens of experience, as noted, yielding ideas such as those which can be expressed employing phrases like 'how it feels to X', ideas which historically resulted in all of the dualist, hard realist notions with which we are so familiar. The anti-realist accounts, possibly motivated by a reaction to this methodology, seem instead to start with the metaphysical – at least on the grounds that a realist self is rejected – and then assign the phenomenological aspects to the illusion of the self or to generating the illusion of the self, although amongst them the more recent psychological and cognitively-focused accounts (such as the following two to be discussed) again commence from a phenomenological perspective.

This emphasis on phenomenology appears to me to be the stronger starting point of the available choices because whatever our selves and our lives may consist of experience is at

¹⁹² Aaron Henry and Evan Thompson, 'Witnessing from Here: Self-Awareness from a Bodily versus Embodied Perspective' in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), pp. 228-249, (p. 228).

the core of existence. Yet I also think that, given phenomenology's centrality, we need to accept that at least in our human case experience is necessarily had through a physical body in one manner or another, and the input there received affects and is affected by any additional, subsequent, or supplementary mental phenomena.¹⁹³ The reader will have noticed that the limited choice soft realist self-view we have argued for is essentially an embodied account due to its nature as an emergent property, while it still contains many elements of a constitutively bodied perspective. It is a self that depends on the body as it is in the world, yet is nevertheless a self that is something more than 'just' the body. This too seems to dance between definitional boundaries, and its realism will have to be further explored, a task that will concern us in this chapter and the final three that follow it (especially in Chapter 6). Our view, moreover, remains open to a fundamental phenomenological challenge which we will now need to consider at length: that of the experientially based yet purely mental account presented by Galen Strawson's thin self/minimal subject/sesmet.¹⁹⁴ If this thin or minimal self/subject view is correct then there simply is no need for the 'something more' – the deeper realism – that our account calls for. We will begin with a presentation of Strawson's thought on the matter.

Strawson's view is based on the 'Experience/Experiencer Thesis', which states that 'experience is impossible without an experiencer', or, in other words, 'Concretely occurring experience can't possibly exist without a subject of experience existing. If you strip away the subject, you haven't got experience any more.'¹⁹⁵ In an earlier work on the self, his fullest and most foundational treatment of the subject, Strawson states this same point clearly with, 'I think that all subjects of experience are selves, and hence that there is

¹⁹³ A multi-layered model of the mind, complementing the previous chapter's model of behavior generation, will be considered in this chapter and the next one.

¹⁹⁴ Galen Strawson, 'The Minimal Subject' in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), pp. 253-278; Galen Strawson, *Selves: An Essay in Revisionary Metaphysics* (Oxford: Oxford University Press, 2009).

¹⁹⁵ Strawson 2011, *ibid.*, p. 253.

a self wherever there is experience.¹⁹⁶ On this, and for clarity in the below, we should note that in none of his writing on the self does Strawson define ‘experience’ in any way, and so it seems reasonable to take his use of the term in the standard sense, and thus understood the experiencer of the experience for Strawson – that is, the thin self/minimal subject/subject (more on these terms Strawson uses and their interrelation below) – does not need to be self-conscious, nor to exist for more than a very short period of time, nor be of any ethical interest whatsoever; it relates rather to the metaphysics of mind, specifically consciousness. It is a ‘thin subject’ and not a ‘thick subject’ (i.e. whole person) or ‘traditional inner subject’ (i.e. soul). Strawson states that, ‘Thin subjects certainly exist, whatever their ontological category or mode of being, for experience certainly exists, and to speak of a thin subject is (by definition) just to speak in a certain way of a feature of reality that certainly exists given that experience exists.’¹⁹⁷

This stance aligns the thin self (or thin subject¹⁹⁸) just with a single experience, and only for the duration of the experience, time after time, meaning that a thin self/subject is very short lived indeed and furthermore that each of us have many, many thin selves/subjects during the course of our lives or even during the course of a single waking day. This is moreover a view of the self that posits it as something that is purely mental. He writes, ‘However body-aware one is, some such grasp of oneself as single as subject in the narrow sense of “subject”, i.e. as single as subject specifically mentally considered, single just qua mental, seems to be built in to full or express self-consciousness’ and ‘one cannot think of oneself in the distinctively self-conscious way unless one experiences or figures oneself as a *thing* in some sense.’¹⁹⁹ In order to strengthen his case for only the mental nature of the

¹⁹⁶ Strawson 2009, *op. cit.*, p. 108.

¹⁹⁷ Strawson 2011, *op. cit.*, p. 261.

¹⁹⁸ Strawson uses the terms ‘self’ and ‘subject’ interchangeably, and his reason for doing so will become clear once the full picture of his self concept has been presented.

¹⁹⁹ Strawson 2009, *op. cit.*, p. 115, 127, respectively; emphasis in the original.

self (again, ‘single just qua mental’) to be necessary and to discount bodily input, Strawson also adds that, ‘a background *psychic* or mental field of awareness, featuring structures of mood, emotion, character, preferences, and knowledge, can serve the same function [as our permanent background awareness of embodiment] and be no less important.’²⁰⁰ It will be clear how such thoughts brush up against the picture of the self that has been argued for in this study, but it will also be clear how the downplaying of embodiment and embedded issues is a major difference between us.²⁰¹

Nevertheless, staying with Strawson’s thin self/subject in order to further fill out the definitional picture and understand how his terms relate, let us first consider his *sesmet* acronym and see how that connects to (is) his notion of the thin self/minimal subject before then scrutinizing the important point of continuity on such a broken view of the self and how Strawson attempts to answer the inevitable diachronic question that arises. Following on from his discussion of the thin self in his 2009 book Strawson relates how self experience is a matter of thinking of oneself as a single thing ‘when experiencing or figuring oneself specifically as a mental subject... it’s a matter of experiencing oneself as a subject of experience that is a single mental thing.’²⁰² This experiential perspective is said to yield the chain ‘subject of experience-as-single-mental-thing’, which Strawson shortens to ‘*sesmet*’. As *sesmets* necessarily involve a subject of experience, and as there must be experience present for it to be experienced – and indeed for there to even be a subject of the experience in question (experiencer) – it is clear how *sesmets* are thin selves/subjects on his account, but also how they are solely connected with the now, with the experience in question. This may appear to hint that a thin self/subject might possibly be more than a

²⁰⁰ *ibid.*, p. 141; emphasis in the original.

²⁰¹ In fact, Strawson states that he thinks embedded considerations have generally been exaggerated in importance in the literature dealing with them; another aspect of his thought that parts company with the view that has been argued for.

²⁰² Strawson 2009, *ibid.*, p. 206.

sesmet if sesmets were taken as foundational or as parts of a thin self/subject whole – although a sesmet could not be more than a thin self/subject (even if we were to allow, contra Strawson, a single thin self/subject to continue to exist across multiple experiences, for example). Strawson considers such ideas but rejects them, preferring instead to see each experience bringing with it a new thin self (subject)/sesmet, and thus equating the terms fully, though he does grant that there might be multiple thin selves/subjects existing together at one time and each connected with one of whatever overlapping experiences were occurring.²⁰³ Sesmets are also, Strawson writes, ‘synergy subjects’ of brain activity and spatially inner subjects (the physical brain’s role on Strawson’s account will be considered below).²⁰⁴

Strawson intends this view to be a third option in addition to the traditional ‘human as a whole’ view and the ‘persisting inner locus of consciousness’ view,²⁰⁵ and he thinks that it can be a third separate view despite its essentially being an ‘inner locus’ concept due to its multiplicity, to its transpiring in a series of thin selves/subjects (each joined with a specific experience) rather than in a diachronic way. He states, ‘I’m assuming that there are as a matter of fact many temporal breaks in the human process of consciousness, and it follows that human thin subjects are as a matter of fact always short-lived’.²⁰⁶ Such a view of course immediately raises a whole person/identity type of objection which Strawson anticipates and replies to before transitioning to a discussion of continuity issues, points which will be instructive for our purposes in this study and to which we will also add

²⁰³ In such cases there will be a dominant or ‘top’ thin self/subject which the experiencer identifies with if that experiencer is a conscious being such as you or I, remembering that for Strawson self-consciousness is not necessary for there to be experience and hence – and this is necessary – an experiencer of that experience. For Strawson an electron, say, has experience and therefore is an experiencer. (For the curious reader I note here that Strawson does follow this thinking all the way out in his later work and argues for panpsychism – the universal consciousness of all matter –; we will consider his case for such in the next chapter.)

²⁰⁴ Strawson 2009, *ibid.*, pp. 323-324.

²⁰⁵ Here I am paraphrasing Strawson’s labels for the views in question, he also calls them ‘the *thick* conception’ and ‘the *traditional inner* conception’, respectively; the emphases are his.

²⁰⁶ Strawson 2009, *ibid.*, pp. 323-325, quoted section on p. 325.

further objections along similar lines in the following. First though, let us consider the whole person objection that Strawson imagines and his response.

Strawson writes:

I agree with you [the objector], in fact, that Louis [or anyone] is and can only be Louis^H [human being] in so far as the name 'Louis' is assumed to refer to an object with relatively long-term diachronic continuity, because I doubt that there's any other object (such as a traditional inner self) that can rightly be called 'Louis' and that has long-term diachronic continuity... I believe that thin subjects [/selves] are always short-lived in the human case, but also because I take it that there aren't any thin subjects at all in the L-reality [Louis-reality] when Louis isn't conscious. *One can distinguish Louis^S [self/sesmet/thin (synthetic) subject] from Louis^H at any time when Louis is conscious, on my view, just as one can distinguish Tom and Tom's hand; but when Louis isn't conscious there is nothing – no object – that is Louis^S (Louis^S is more like a goose pimple than a hand).*²⁰⁷

Here we can see the extent of the radicality in which Strawson reduces the self yet still claims there to be a self in an 'inner subject' manner, albeit never a single (enduring) inner subject. In some ways this passage reveals similarities with Hume's account as explored earlier, in which there is simply nothing about the self, or to the self, when one is not conscious; this moreover further highlights the affinities that Strawson's account has with other anti-realist views. Yet what is also noticeable is that by bringing in analogies such as a hand or a goose pimple there is a sense here in which the self might be considered by Strawson to be 'owned' in some way, perhaps by the whole person (e.g. Louis^H), and

²⁰⁷ *ibid.*, p. 335; emphasis added.

perhaps even without Strawson's noticing it – as a background or assumed (*pre*-sumed, unexamined) aspect within his thought), taking us back to the ownership issues we found with the soft anti-realist views explored in Chapter 1, and in particular regarding Narveson's ideas. It may be that the metaphors are simply misleading, but by making the comparisons that he does Strawson invites us to ask what exactly Louis^H is for Strawson vis-à-vis Louis^S such that Louis^H might relate to Louis^S as Tom does to his hand. Does it matter that the identity of Louis^H appears to necessarily entail further facts beyond what Louis^S can give? If so, how? Additionally, what happens to Louis^H on the L-reality when Louis is not conscious and there is no Louis^S? We will return to such thoughts on bodily and psychological issues in our objections to Strawson's account and so at this juncture I wish only to highlight the closeness of this view with those we earlier found to be problematic and to suggest that Strawson's self concept might not be as much of a third option as it appears given its crossovers with other soft anti-realist accounts.

In responding to this imagined whole person objection Strawson also raises the important point of continuity on his self concept, to which we will now turn. Strawson writes that, 'This succession of thin subjects is, I suggest, the reality that underlies any experience of being a persisting inner self.'²⁰⁸ It would seem by this that the parade of thin selves/subjects – the continuous flow of one single and distinct thin self after another – generates for a being with awareness the perception of having an ongoing self. Is this enough to count as continuity? Is it sufficient to give rise to the kind of (often strongly held) 'inner locus' stance that many of us have as our default or unreflected position? The questions will invite their own answers, and, when viewed through the lens of his overall conception, one is tempted to even see something of a realist self here in Strawson's account, perhaps a type of 'internal voice' or 'me-ness' that arises naturally out of this

²⁰⁸ *ibid.*, p. 337.

succession, particularly given that Strawson considers thin selves/subjects to be both metaphysically and physically real objects. Yet on the other hand, as stated above, Strawson himself directly rejects any realism to the self, and it is noteworthy that he does not actually describe the self as being ‘persisting’, but rather only the experience of it as being so. Perhaps then we must either call this a soft anti-realist view or a quasi soft-realist view, depending on which side of the nuances involved we wish to favor. Nevertheless, in returning to our above questions, we still must see how, given a continuing succession of ultra short-lived selves which may or may not overlap, continuity is meant to work, or to be generated or to be defined, on this account.

As remarked, Strawson’s self concept includes a Parfit style start-stop-start²⁰⁹ identity as during unnoticed gaps there simply is no subject; again, ‘a normal waking day involves the existence of many thin or minimal subjects’.²¹⁰ If we are concerned to find any realism at all to Strawson’s self then this lack of diachronicity argues against us; and without diachronicity what kind of continuity is actually taking place? Though we might pause to ask a deeper question: Is any continuity really necessary? After all, if there is no self then its ongoing-ness or lack thereof is of no matter. Still, there is the phenomenological aspect of the ‘experience of being a persisting inner self’ that requires attention, and in working through that Strawson states that his concern ‘with the self is simply a concern with the subject of experience conceived as something that isn’t the same thing as a whole human being’ and hence the reason he prefers to use the ‘less exigent’ term *subject* over *self*.²¹¹ This is a subtlety worth taking note of, for though the term ‘subject’ might admittedly require less conceptual substance to it – at least in its common usage – we still find ourselves in need of an explanation for that undeniable experience of continuity one has as

²⁰⁹ See Chapter 1 Section 1.1.B (‘Four accounts of the self: The hard anti-realist position’) for details.

²¹⁰ Strawson 2011, *op. cit.*, p. 262.

²¹¹ Strawson 2009, *op. cit.*, p. 345.

one moves from day to day. A further element to point out here is that the use of ‘subject’ in this way by Strawson allows one to conjecture that his account might match well with how other researchers have used the term ‘minimal self’, which does include a notion of ownership and agency in a ‘pre-reflective consciousness of oneself as an immediate, embodied subject of experience’,²¹² also corresponding to Rochat’s ‘ecological self’ (first credited to Ulric Neisser) as described in the above.²¹³ Is a thin self/subject essentially the same as these other thinkers’ minimal selves? If so, would that explain the experience of continuity (say through its elements of ownership and agency)? Let us continue exploring this question for Strawson’s account has more surprises in store.

As to that appearance of continuity that we typically (phenomenologically) experience, Strawson explains, ‘A further reason why the appearance of continuity is unsurprising, on his view [William James’] as on mine, is that the subjects – subjects/experiences – in question arise successively, in a single person’s brain, from brain conditions that have considerable similarity from moment to moment even as they change’,²¹⁴ and that ‘The “stitching software” that underwrites our sense of being a single persisting subject – and delivers a sense of the flowing continuity of experience (for those who have such experience) – is as remarked extremely powerful.’²¹⁵ This is an appeal by Strawson both to the phenomenology of the self as illusory and to the physical brain as paramount, once more reopening previously considered issues, this time regarding the brain/mind relationship and also appearing strikingly similar to the kind of ‘self as brain’ neurophysical arguments that Strawson himself explicitly rejects when building the case

²¹² Vogeley and Gallagher, *op. cit.*, pp. 118-119.

²¹³ Rochat, *op. cit.*; Ulric Neisser, ‘Five Kinds of Self-Knowledge’, *Philosophical Psychology*, 1 (1988), 35-59; see Chapter 2 Section 2.3 (‘Bodily and embodied issues’).

²¹⁴ Strawson 2009, *op. cit.*, p. 353; James’ referenced work is his *The Principles of Psychology*, 2 vols. (New York: Dover Books, 1890/1950).

²¹⁵ *ibid.*, p. 355.

for his self concept (labeling them ‘brain-system’ arguments).²¹⁶ Due to this earlier rejection Strawson cannot mean to equate a self with the brain, and so perhaps he wishes instead to equate an *impression* of a self with the brain; such would make much more sense as for Strawson there is anyway no self to speak of. Yet if this is what he means then he still has to *explain* continuity. Are similar brain conditions sufficient to yield a feeling of an unbroken self? What is this ‘stitching software’, and if there is no self then why does it exist? For evolutionarily valid reasons? What might those be? What could be the biological advantages of so deceiving ourselves? Would we not survive better if we all knew that there is no actual self but only successions of thin selves? If not, why not? In which instance would we be psychologically healthier? Why? Given the effort he has put into arguing for his self-view it would seem that Strawson must think we would be better off by accepting his position. Yet if that is the case then why did nature apparently steer us wrong? There might be reasons for this, and possibly very good ones, but we would like to at least have some suggested. Merely remarking that it is ‘unsurprising’ does not account for any of these subsequent queries even should we agree that the experience of the self is at core an illusion and that all that is ‘really there’ is a string of thin selves/subjects in times of awareness and nothing at all otherwise. Continuity is also of course unsurprising for many other self accounts, not least traditional (and non-traditional) realist ones which thereby do provide an explanation for it: namely that there actually is a continuing self (of whatever kind). For realist views the explanatory burden is much less since the surface matches the depths: we experience a self because there is one. On the thin self/subject view there is no such match and so the burden is greater, and that has not been met; here we still need a proper response to the challenge presented by our experience, one that not only accounts for it but that gives a reason why we should have it if that experience is grounded in nothing but a mirage.

²¹⁶ *ibid.*

We will return shortly to brain-empirical issues on this account, but first on the physical broadly considered Strawson does raise a very good challenge to our typical perceptions (and judgments) when he asks:

We take it that we detect a straightforwardly objective, theory-independent feature of physical reality when we judge that it comes in very small units spatially speaking. Why should we think that this is unlikely to be true temporally speaking? Are we going to quantize everything else in nature and leave selves and experience – the real physical phenomenon of consciousness – as densely continuous?²¹⁷

This is a fair question, to which it is perhaps best replied that how we *experience ourselves* – as individual and cohesive selves – is as being densely continuous, even if we accept that there are gaps in one’s consciousness.²¹⁸ If we do agree to ‘quantize’ the self in the manner Strawson suggests then we inevitably run into the same problem of overlapping experiences/unclear divisions between experiences and the multiple simultaneous selves/subjects that such must give rise to by the very definitions and parameters of Strawson’s account, a feature which Strawson has admitted to in his stipulation that in those situations there will be a dominant or ‘top’ self amongst the concurrently existing manifold selves (e.g. co-existing due to each being generated by/associated with a simultaneously occurring experience; see footnote 202 and the accompanying text above). This ‘top subject’ is thereby the presently experienced self/subject of experience, the self that we feel to be ‘us’.²¹⁹

²¹⁷ *ibid.*, p. 392.

²¹⁸ The reader will recall that this study argues against such gaps, that indeed consciousness does not cease throughout a life, although awareness does. This topic will be returned to in the next two chapters.

²¹⁹ Strawson 2009, *ibid.*

Moreover, although Strawson accepts that such multitudinously contemporary selves/subjects might be the case of the matter, for him it is not so much a theoretical problem as a question of fact (i.e. empirically so), and he thinks that what scientific-psychological investigations reveal about the area may catch us off guard. I do not find either of these positions (the ‘top subject’ and the ‘wait and see what science says’ positions) to be especially strong however, as in an account that – rightly, I think – wishes to anchor its self concept phenomenologically simply stating that we may leave it to the hard sciences is setting the question aside a little too easily, and furthermore that by admitting a grouping of multiple simultaneously occurring ‘selves’ – selves that are defined as ‘subjects of experience’ – wherein one is ‘dominant’ and therefore the phenomenologically felt or expressed ‘subject of experience’ there hardly seems to be any need (other than theoretical/definitional) for the other sub-existing and non-experienced selves. Apart from their fitting the thin self/subject structural model just what work are such ‘selves’ doing? They are simply markers attached to other and – apparently – unnoticed experiences. This raises the further phenomenological question of what an unnoticed, *unexperienced* experience is not only meant to do but meant to be. If it is something that could be set aside (‘bracketed’²²⁰) and examined, then perhaps we could say that it is at that point of analysis, and only at that point, when the experience in question actually gives rise to a thin self/subject. If so, however, then we would have an account wherein thin selves/subjects pop into and out of existence solely with each noticed/felt experience, and since we can purportedly have unnoticed and non-felt experiences each experience would not have, nor have need of, an associated thin self/subject. (I.e. experiencer; and note that this would violate Strawson’s core stance of each experience

²²⁰ See Section 3.3.C below (‘Phenomenology 3: Going back to Husserl’) for an overview of phenomenological methodology.

necessarily having an experiencer – unless ‘experience’ also meant ‘stimulus’ in some sense; possibly as a unit of preconsciously processed brain data?) Perhaps Strawson would be happy to accept this adjustment, but he has argued otherwise and we can do no more than to deal with what we have.

Finally, let us expound on the role that the physical brain plays in Strawson’s thin self/minimal subject account. Towards the end of his dense and highly abstract book focused almost entirely on metaphysical issues concerning the self, Strawson seems to make an abrupt empirical turn. Without going into too much detail we may note as background that Strawson favors a view of reality that is fully physical, that posits that there are only physical phenomena and that all mental (or otherwise traditionally thought of as nonphysical) phenomena are in fact physical at core, although perhaps in a way that is simply not properly understood at present – at least by the mainstream perspectives.²²¹ This alleged empirical turn, if such it be, entails Strawson shifting the gears of his running argument to the effect that he suddenly drops phenomenology as the guiding foundation and states that science:

will, I venture, show that a radical change of subject matter requires a new thin subject, given that a thin subject is a neural synergy of a certain particular sort. On this view, it’s a fact about brains that new content of a sort that amounts to a strong content break necessarily involves a new seizing, a new synergism, and hence, *eo ipso*, a new thin subject.²²²

This is curious. Are we to start thinking of thin selves/subjects as ‘neural synergies’? What

²²¹ Strawson 2009, *ibid.*, see in particular Part 6 (‘Metaphysics: Preliminaries’). In the next chapter of the present study we will look at length into Strawson’s panpsychism, which he bases on his physicalism.

²²² *ibid.*, p. 399.

might such be? Given the context of the scientific investigation of the brain, perhaps an image occurs to us of something like a grouping of neurons coming together at one part of the brain and then another as this experience turns into that one. Strawson goes on:

The passing thoughts or experiences or thin-subject selves are wholly neural phenomena, neural synergy phenomena, and the brain also provides a place of residence for everything that is needed for Louis' experience to be exactly as it is compatibly with the fact that Louis' existence as a subject of experience...consists in nothing other than the existence of the short-lived thin subject selves I have described.²²³

He adds:

each such experience-and-subject being a primary unity, a matter of a certain sort of upsurging of activity in and across neurons, each such upsurging effectively numerically distinct from the next... I think that this is what the phenomenon of there being a subject of experience actually consists in, in the human case. This is the reality that underlies all the subjective phenomena of continuity and flow in experience, such as they are, and the whole natural picture of the persisting inner self or subject.²²⁴

Again, this appears to look very much like the modular accounts of the self mentioned above whereby the brain has a specific area(s) whose functioning yields the feeling of being a self, the very accounts which Strawson himself denies outright (calling such 'brain-system' arguments, see the text to footnote 215 in the preceding). How Strawson's view differs from such concepts apparently lies in the fact that Strawson would not define a

²²³ *ibid.*, p. 404.

²²⁴ *ibid.*, pp. 413-414.

specific intracranial location, writing ‘On this view, there’s simply no locus in the brain, however scattered, that is (a) the locus of the subject of experience and (b) distinct from the place where the neuronal activity in virtue of which the experience has the content it does is located.’²²⁵ We must conclude, it seems, that although the thin self/subject as ‘experience-and-subject’ must somehow be occurring within, anchored to, and compositionally structured by the brain (being ‘neural synergies’), it is not thereby only related to one area of the brain but can be found anywhere as varying experiences stimulate different parts of the brain and generate thin selves/subjects-qua-neural synergies hither and thither. All that really appears unique here is the insistence on ‘no locus...however scattered’; otherwise it simply is not only a ‘brain-system’ account (to use Strawson’s term for such conceptual frameworks), it is the same ‘brain-system’ account that we have already encountered. (Moreover, if we factor in Gazzaniga’s ‘constellation of consciousness’ exposition we see that there is really no single locus strictly required on the modular view either.²²⁶ We might also wonder, given basic biological facts about cell structure and function, whether the entire brain really is flexible enough in all of its many parts to allow such a scattered ‘neural synergy’ explanation as Strawson’s to occur.) We still have absolutely everything about the self (or subject) on this view as being fully and exclusively brain related, and what is more, fully and exclusively *physically* brain related neurally; any room for psychological mind aspects of the self – to say nothing of other aspects such as the emotions, intuitions, or consciousness considered in its complete and emergent senses – has been done away with. When all is said and done we are left with nothing but flashes of ‘neural synergies’ that align with specific and brief experiences in

²²⁵ *ibid.*, p. 416.

²²⁶ The reader will recall that Gazzaniga’s proposed ‘interpreter module’ was just that – at present a proposition. No empirical evidence has yet been found for such, though Gazzaniga does hold that there are theoretical reasons for thinking it may be there. If it were to be located and identified then I suppose that it would qualify as a locus in the sense that Strawson is using here. Though if that were the case, if science did discover a specifically placed ‘interpreter module’, then one might think that Strawson would accept such given his above comments on what science may show. See Section 1.1.A of Chapter 1 for Gazzaniga’s comments (‘Four accounts of the self: The soft anti-realist position’).

the world.

When considered on the whole then, the objection to our self concept that Strawson's account seems to present most forcefully is this: that experience – along with its attendant 'experiencer' (the thin self/minimal subject-cum-neural synergy) – is by itself enough to account for all that is necessary for the self, nothing more beyond it need be assigned. We do not require a real or lasting self at all to have the kind of experiential lives that we enjoy on a day-to-day basis, including, it would seem, our emotional lives; we simply require a thin self/minimal subject. There is some weight to this picture, and we must admit that what experiences we have, be they sensual, psychological, emotional, intuitive, or other, do and must involve our physical brains. Emotions, for example, are brain-generated reactions to stimuli received via the senses (naturally including perception), and as such both the input and output sides have been processed by the brain and then expressed – in whatever form – via the brain through its directives. In these ways the self is certainly brain-based as a (normally) living human being is brain-based.²²⁷ Does this though mean that our experiences, and hence lives, are exclusively brain-based? Fully and only ever physical? At root this comes down to the position that we wish to take on the definition and role of consciousness, to whether or not we wish to embrace the kind of reductionist picture that Strawson gives here, or rather a looser but more expansive picture such as one that Nagel might argue for.²²⁸ Is a human life 'just' a physical machine moving in and through its environment? Is any life 'just' that? How much more or less is involved for a human being alive in the world than for a dog? Or an ant? These issues of consciousness are highly interesting and deeply entwined with our topic of the self, and they will be repeatedly

²²⁷ Comas present interesting and challenging cases here, and there are many debates as to how life ought to be defined in such instances, particularly when the brain is dead but the body is kept alive. Interested readers might find the following collection a helpful way into some of the issues involved: *The Cambridge Companion to Life and Death*, ed. by Steven Luper (New York: Cambridge University Press, 2014).

²²⁸ E.g. Nagel 2012, *op. cit.*

discussed in the remainder. For the moment however I wish only to indicate that the stance which will be adopted in this study is that although there are different levels to consciousness it involves emergent properties which cannot be accounted for simply by the summing up of parts. This is not to claim that consciousness itself is an emergent property – although it very well may be one – but instead to state that when it comes to the experiences connected with human existence there is more at play (and certainly more internally felt) than descriptions of neuronal networks appear able to account for, and that is so because of the top-down manner of our approach to the collaboration of our parts. This approach, this feeling, could of course be leading us astray, and the physical side might be enough – at least explanatorily; that door should for now be left open. In the meanwhile we will simply have to keep these difficulties in mind as we continue.

Our agnosticism notwithstanding, in considering Strawson's self-view from this physical and neuronal vantage point, as well as from its more ontological one-self-per-experience theoretical perspective, we see that it does not appear to give us the full picture of human experience due mainly to its reductivism. This is a judgment that needs to be argued for, and it will be: in the below we will assert that in addition to the objections already given to Strawson's account we may add that when a broader life is analyzed – and not simply a moment of experience, or even a string of such moments: that is, when granting a full phenomenological basis to our considerations – we are left with a view of the self that is a soft realist one. The key term here being realist. In presenting our case we will primarily use evidence from other phenomenological researchers lest our conclusions be dismissed as outside the scope or irrelevant to the challenge as it stands.

To undertake these considerations of life on a grander scale let us think about the 'paradox of human subjectivity' as described by Dorothee Legrand: one's body is experienced as

one's own belonging to the world it inhabits and yet it cannot be an object in the world, it must 'encompass a lived subjectivity', part of which is being able to experience the world's objects.²²⁹ What this means is that although the body is ostensibly an object in the world it is not an object *to me* as long as it is *my* body; I cannot experience it in the way I can other unrelated objects. This does not preclude reflection on one's body, but it does mean that 'The seeing body is thus *transparent* in the sense that one experiences the world *through* it...the bodily self experiences itself as a volume localized in space and oriented relative to things in the world.'²³⁰ Legrand points out that for both Maurice Merleau-Ponty and Husserl consciousness of the world and self-consciousness are contemporaneous, synonymous;²³¹ again, the notion of the world as experienced through 'me' through the body. This in itself does not discount the thin self/minimal subject as an explanatory device, but it does highlight an important aspect of experience that Strawson's self concept is missing: namely the idea of placement and relation, that experience is more than just received, it is always interactive and never purely passive.

Strawson's account in his *Selves* book, and all of his later defenses of it, are written consistently and exclusively from the perspective of an experience requiring an experiencer, thus placing experience itself foundationally. Yet we might query how this could be without resorting to some kind of infinite regress? How could an initial experience grant a first experiencer? That is, in the absence of an experiencer how could there ever be an experience? Such questions are never asked in Strawson's works, but in posing them we see that Strawson is forced to hold the self – albeit his specific thin self/minimal subject version – as being something antecedent that must be found by working backwards from

²²⁹ Dorothee Legrand, 'Phenomenological Dimensions of Bodily Self-Consciousness' in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), pp. 204-227 (209).

²³⁰ *ibid.*, p. 215 and p. 217; emphases in the original.

²³¹ *ibid.*

the starting point of an experience as it has been experienced. If so then experience is not the foundation for a self, a self is the foundation for experience (working backwards), and this necessitates some substance to the self that goes beyond what a thin self could provide for it because that same substance must be there *in the absence of experience* in order for that experience to be experienced in the first place. This is a realist stance, and clearly it is not Strawson's thin self/minimal subject, not by a long shot. Methodologically this going in reverse is sound from a phenomenological perspective, but it does not lead to Strawson's conclusions.

Strawson moreover seems to completely cut off the possibility that the self could be something that is emergent and developing, and although I think he is right to do so given his convictions regarding one-experience-one-self, I do not agree with those convictions. On a view such as his the self cannot possibly be thought to develop in any way whatsoever as it lacks real diachronicity, yet without diachronicity we fall into the regress trap just mentioned once we look hard enough at the whole picture and not only at the momentary one. Purely procedurally speaking, starting from experience and then looking for the self/subject (experiencer) of it does work and can lead to Strawson's conclusions, but only if we ignore the regress questions it raises. Once we ask such we find that there must be some self initially there, but if each self must have an experience, and each experience must have a self, we find ourselves circling and circling with neither side of the equation really capable of getting the whole process initiated. What I think is being overlooked by Strawson is that we are not merely subjects *of* experience; in an important sense we *are* experience, in an ongoing and cumulative manner as the two are built together and reciprocal – but that constructive process must begin from somewhere. That 'somewhere', we have argued, is a (soft) realist self.

This issue also highlights an aspect of experience that Kristjánsson's self concept on its own is missing, and hence the need to expand and refine that view as the previous chapter sought to do.²³² What both Strawson's and Kristjánsson's accounts fail to take note of is this bodily self-consciousness as a necessary part of the self and any experience the self has; it is the preconscious awareness of the bodily-inhabited subjective sense. When bodily self-consciousness is considered it leads to the realization that any notion of 'me' has to include the idea that the 'I' cannot be separated from the body and will always experience the world via, and in connection to, that body. This conclusion must be arrived at due to its quality of being preconscious: it is a sense, an awareness, that one cannot shake, that is part and parcel of what it means to be a human animal. It is in our DNA, it is how we are built, it is a biological fact of nature. We can conceivably try and talk ourselves out of it with our rational minds (Descartes of course providing a very skillfully done example of this), but as soon as we stop convincing ourselves and return to life as it is lived that sense comes roaring back, expressing itself through our behavior, reactions, intuitions, judgments, emotions. Vogeley and Gallagher's use of 'minimal self' (discussed in the previous chapter), with its pre-reflective consciousness of embodiment, comes close to factoring in the importance of bodily self-consciousness, but to me it does not take the concept deep enough, does not give the body enough weight; Strawson's is shallower and lighter still.²³³ To repeat, as counterintuitive as it might seem to put it this way, we 'have' the body that we 'are'. (Note here too that the difference between a 'pre-reflective consciousness' and a 'preconscious awareness' is not merely cosmetic, it is one of degree and to equate the two would be a conceptual error. A preconscious awareness indicates an awareness not only before one stops to think about it but in a way almost beyond how one can even think

²³² Compare our self-view with Kristjánsson's bare account in Chapter 1, Section 1.1.C ('Four accounts of the self: The soft realist position').

²³³ Compare the discussion in Chapter 2: Section 2.3 ('Bodily and embodied issues') with Strawson's account here.

about it; it is ‘in the bones’, as it were.)

Legrand summarizes the dual role of subjectivity and intentionality that the body plays with the following: ‘the physical body can be taken as intentional object and be experienced as expressing its subjectivity...it can also be taken as intentional object and be experienced as fully belonging to the physical world in which it is embedded’.²³⁴ This is the ‘paradox of human subjectivity’ referred to earlier: we can experience our body as ourselves – whatever we take the self to be –, as fully identified with the ‘me’ that we contemplatively regard or that we reflexively enact, as that point of intersection with the world. Or we can take the body to be something we ‘own’ and think about and treat it the way we do other objects. In either case, however, the body is always and must be lived, inhabited, *embodied*, and hence it can never truly be ‘just’ an object in the world. Understanding the body in this way brings us closer to the self concept that has been argued for in the preceding (with its bodily presence obligation), and it indicates that we at least require – simply to account for experience and nothing more – an explanation of the relationship between the sense of being an experiencer and the body through which that experience comes. Strawson does not offer such an explanation as he does not consider the body to be an important element involved in the self, nor indeed does he seem to consider the body to be a necessary element.²³⁵ (Though to borrow from Strawson’s style we might try to put the idea in play here as something like ‘the experiencer-as-mental-as-body’.) We might also point out that this explanatory requirement that is included by inserting bodily self-consciousness into the phenomenological picture – which again we must do if we are to really reflect on human phenomenology – remains in place whether or not one accepts the argument for the body both giving rise to and being dependently inhabited by the self.

²³⁴ Legrand, *ibid.*, pp. 225-226.

²³⁵ Strawson’s metaphysics, to be discussed in the next chapter, allows for even subatomic particles to have (to be?) thin selves/minimal subjects, and he remarks that he takes this to be the case.

Thus for bodily issues; our next avenue of concern in responding to the thin self/minimal subject also covers points referred to above from a phenomenological perspective, namely, psychological and cognitive issues. Dan Zahavi, also citing Husserl's view, states that 'If we take a closer look at Husserl's analysis, we will somewhat surprisingly find him stating that the ego is present *everywhere* in the living present, and that even the anonymous stream of consciousness would be unthinkable without an original ego-pole as the center of action and affection.'²³⁶ This notion of an 'ego-pole' might be considered as another step towards a more holistic understanding of the self: the 'axis' that ties together both one's automatic affective nature and one's active nature (whether active in the sense of intuitively reactive or as rationally chosen); the self must be both preconscious and reflectively definable. Zahavi lends further support to this type of self-view when he writes that:

Contrary to a widespread misunderstanding, the notion of a transcendental ego is not bound up with an idea of an autonomous free-standing ego. To defend the existence of a transcendental ego is to be committed to the view that the first-person perspective is a necessary condition of possibility for manifestation. It neither commits one to the idea that it is a sufficient condition of possibility, nor does it necessarily involve a failure to recognize the role of passivity.²³⁷

Zahavi seems to be arguing that an ego that is transcendent in this way could still be just as physically bound (in this case psychologically so, that is, as being dependent upon a functioning physical brain) as any other element of the creature might be without perforce

²³⁶ Dan Zahavi, 'Unity of Consciousness and the Problem of Self' in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), pp. 316-335 (p. 322); emphasis in the original.

²³⁷ *ibid.*, footnote 2 on p. 325.

adding anything more. What is only implied, however, is the presence of consciousness, for how could a first-person perspective be in place without consciousness? Once more we see that any consideration of the self seems to very quickly become more complicated than we wish it to be, and although consciousness issues will be treated to a more in-depth discussion in the next two chapters than they have heretofore been, we can for the moment remember that along with bodily presence our definition of the self does call for consciousness as well. That it does so agrees with the necessity of a first-person perspective, although whether that is in its favor or not remains an open question as the more factors a definition has the more points there are for attacks to be made on it.

Moving further along these lines and recalling what the previous chapter attempted to argue, when the self is understood to be bodily emergent we actually can have our (biological) cake and eat it too. Moreover, and once more contra to the thin self/minimal subject as a sufficient explanatory concept, when we stop looking only at one experience in isolation and move to a continuing series of experiences it is clear, Zahavi relates, that each new encounter does not have its own ‘for-me-ness or mineness’.²³⁸ Any number of incidents might not be noticed at all (in awareness) and yet still flow from and be responded to by – in whatever way – an ego (or ‘ego-pole’) point of view.

Zahavi however does note that, ‘On this view, there is no pure experience-independent self. The self is the very subjectivity of experience and not something that exists independently of the experiential flow.’²³⁹ The idea of the self as the subjectivity involved in the flow of experience is perhaps sufficiently vague that it could be interpreted to fit a number of rival accounts (maybe especially non-realist ones), although Zahavi does also claim that the

²³⁸ *ibid.*, p. 327.

²³⁹ *ibid.*, p. 328.

diachronic unity of the self is not threatened by interruptions to consciousness (e.g. dreamless sleep, being comatose, blackouts, et cetera) since the self's identity stems from its givenness to the organism and not its temporal continuity.²⁴⁰ In these ways Zahavi's Husserlean-based ideas grant more substance to the self than Strawson's do, yet still not as much as this study has argued for (importantly leaving out the emotions). It may be interesting to pause here to ask whether on our view there could be an experience independent self. In response we would need to say that yes there could, for our definitional elements do not require ongoing experience per se, only the emotive sets, consciousness, and bodily presence, none of which demand active (external) occurrences, although each will often be informed by such. Regardless of however much we agree or disagree with Zahavi, what is pertinent at this point is how his arguments indicate further insufficiencies with Strawson's minimalist self-view.

Strawson does state that 'the thin notion of the subject doesn't exclude the use of other notions',²⁴¹ and it may be that he means his view simply as a way by which experience can be accounted for,²⁴² but if that is the case then his self theory tells us little and leaves out much, especially when we step back to consider the phenomenology involved in a complete life and not only one instant within that life. Moreover, if we do merely have a start-stop-start subjectivity/selfhood, what kind of substantive – that is, beyond the moment – subjectivity or self can we be said to have? For Strawson and others that of course begs the question, and the ready reply is that we have no lasting subjectivity and no self at all, plain and simple. Yet such a position raises the same problems that we have been dealing with all along: considering a whole life, a life with a body and at the very least the

²⁴⁰ Zahavi, unlike Strawson, does seem to grant diachronicity as important and the self as unified.

²⁴¹ Strawson 2011, *op. cit.*, p. 262.

²⁴² Strawson (2011) refers the reader to parts 6-8 (titled: 'Metaphysics: Preliminaries', 'Metaphysics: The Question of Fact, 1', and 'Metaphysics: The Question of Fact, 2') in his 2009 (*op. cit.*) and states that he holds that there is no distinction between the existence of a subject at any time and the existence of its propertiedness, part of which is at least the subject's experience at the time; see p. 276 in *ibid.*

appearance – if not the actuality – of a unity of consciousness, both of which feed into and are fed by the experiences we encounter, merely transitory subjectivity seems insufficient. We cannot of course go from this doubt straight to a realist self, however, nor can we consider the objections that could be made to a fuller and bodied (soft) realist conception of the self to have all been presented by Strawson’s account, but we can admit that the picture is slowly getting clearer and that issues related to consciousness appear to be central and to be crucial; such will be wrestled with in Chapters 4 and 5. In this chapter, however, we will maintain our focus on phenomenological issues as we consider another recent and popular self concept from those quarters: Barry Dainton’s stream of consciousness-centered phenomenal self.

B. Phenomenology 2: Dainton’s phenomenal self

Dainton’s self concept is based on what he terms ‘extended C-systems’ (‘C’ for consciousness), which are by definition more or less a mind, but a mind specifically understood as an entity ‘whose persistence conditions are defined in phenomenal rather than psychological terms.’²⁴³ What this means in practice is the viewing of a person’s mind as consisting of experienced consciousness states along with ‘causal-functional P-relations’ (psychological relatedness), which are items that string together the parts of the mind where consciousness, or the ability to produce a stream of consciousness (‘C-relatedness’), does not reach. (These may or may not be similar to what Strawson referred to as ‘stitching software’, Dainton does not comment on the issue.) What is most essential on this theoretical standing are those streams of consciousness – one’s singular stream of consciousness as one experiences it throughout one’s life – but since that alone cannot account for the totality of the mind Dainton also brings the notion of the assisting

²⁴³ Barry Dainton, *The Phenomenal Self* (Oxford: Oxford University Press, 2008), p. 177.

P-relations into his position. This differs from a view like Parfit's,²⁴⁴ which is also based on psychological continuity and/or connectedness, in its perspective on continuance. To illustrate his understanding Dainton gives the example of Sam, whom if we destroy but then a second later bring New Sam into being, would be treated as two numerically distinct individuals on so-called 'P-theories' (that is, psychological connectedness theories such as Parfit's), yet 'for the C-theorist the lack of causal dependencies in cases such as this is quite irrelevant; all that matters is whether Original Sam's experiential powers are C-related to those of New Sam. Could these powers produce co-conscious experience? If they could then we are dealing with a single persisting subject'.²⁴⁵ (It should be noted that 'co-consciousness' is used somewhat specially by Dainton to indicate phenomenological unity within a single conscious state, both in synchronic and diachronic forms.²⁴⁶) What is here essential to the self – and as with Strawson both 'self' and 'subject' are used by Dainton interchangeably – is quite simply that stream of consciousness and its linkages.

Dainton argues for his self-view via a series of Parfit-style thought experiments, using one in particular (based on Robert Nozick's well known 'experience machine'²⁴⁷), the VR-4 device, that provides its user with '*virtual reality of the fourth degree*...the VR-4 program does not install a new psychology into your brain, it completely bypasses your own psychology, and furnishes you with an uninterrupted flow of experience that is entirely computer-driven.'²⁴⁸ Since by this machine a person loses all psychological connection with themselves prior to plugging into it yet still, while plugged in, feels that they are

²⁴⁴ For a summary see again Chapter 1 Section 1.1.B ('Four accounts of the self: The hard anti-realist position').

²⁴⁵ Dainton, *ibid.*, p. 179.

²⁴⁶ Dainton helpfully provides a short glossary of technical terms at the back of his book, the definition there for 'co-conscious' reads as follows: 'experiences are co-conscious if they are phenomenally unified, i.e. experienced together as parts of a single conscious state; co-consciousness comes in synchronic and diachronic forms: the former holds at a time, the latter over short periods of time (within the confines of specious presents)'; see *ibid.*, p. 423.

²⁴⁷ Nozick wrote about this in his *Anarchy, State, and Utopia* (New York: Basic Books, 1974).

²⁴⁸ Dainton, *op. cit.*, p. 18; emphasis in the original.

experiencing everything that occurs as happening to oneself and indeed as never losing track of remaining oneself, Dainton thinks that the intuitive judgments reached via such considerations strongly suggest that only our stream of consciousness experiences are what is needed for our selves. In the human case this means the aforementioned ‘extended C-system’, yet in its most essential case it need not even be extended, say in a creature that has no psychological dimension and is purely phenomenal.²⁴⁹ Summarizing, Dainton writes that ‘According to the C-theory, you are identical with a future self or subject if you are connected to the latter by an uninterrupted potentiality for experience; there is no restriction on the sorts of experience this future self can have, or their physical or psychological traits.’²⁵⁰

It can already be seen that we will have similar objections to this account as we had to Strawson’s self concept, but let us first examine the place of the body here, for that is something that Dainton does consider, before we then take note of the irrelevance – and the importance of that irrelevance – of affection and intuition on Dainton’s view, as well as the other psychological aspects that are downplayed and grouped together by Dainton due to their absence of experiential manifestation.²⁵¹

The body, to begin with, is dismissed by Dainton when he asks whether the effects of body chemistry on one’s mental experiences should be considered or not as elements of one’s C-systems (= self) and then concludes that such should rather be seen as part of that in which the C-systems happen to be based (but, it would seem, need not necessarily be; again, the entire focus is on consciousness viewed phenomenally). This stance in itself

²⁴⁹ *ibid.*; see the discussion on pp. 188-192 and especially p. 190.

²⁵⁰ *ibid.*, pp. 195-196.

²⁵¹ Of this latter category Dainton lists as examples, ‘Non-repressed propositional attitudes, many of the more complex emotions, certain character traits, and mental abilities’; *ibid.*, p. 174.

strikes me as at least a flirtation with dualism or perhaps the possibility of dualism, and Dainton does indeed take the notion of a mind/body split – or maybe better put as a body subtraction – further by suggesting that ‘Perhaps the best course [in considering the role of the body] is to accept a certain blurring of boundaries [i.e. between mind and body]’,²⁵² a statement which he follows with an argument propounding that such bodily effects (i.e. of body chemistry) are really nothing to be concerned with, for an envatted brain with the right equipment hooked up to it would still let ‘me’ (C-system self) have the same experiences as I would in my normal embodied or physical or day-to-day state.²⁵³ We have in the above already considered the case of an envatted brain and need not return to it in reply;²⁵⁴ what is most pertinent here is instead the willingness Dainton displays to entirely eliminate bodily concerns on his self-view. Since his account is based on, and consists in, one’s/the/a stream of consciousness perhaps we should not be surprised by his lack of concern for the body, but Dainton nevertheless feels the point worth pressing and continues his reasoning with more thought experiments meant to demonstrate the lack of necessity of a body while maintaining a fully-fledged self.

Thus it is that a little later in his work Dainton considers a situation whereby nanorobots sever all of the neural connections between a brain and body and then restore those same connections, finding that the related phenomenology of such a series indicates that the self’s embodiment depends on, 1) sensory links, 2) action links, and 3) corporeal form. While this might be adequately representative as far as a description goes, what is telling is that throughout the passage phrases such as ‘your/our body(ies)’ are used, an indication –

²⁵² *ibid.*, p. 202.

²⁵³ *ibid.*, p. 203.

²⁵⁴ As a reminder of our earlier arguments, the absence of perception and action and the entirely unclear situation regarding thought and emotion in such cases were found most troublesome: even if the ‘right equipment’ were able to provide perception of some sort the brain would still be unable to act and the degree and type of thought and emotion that such a brain might encounter or (attempt to) generate cannot be adequately approached in any way save by a fairly uninformed speculation; see Chapter 2 Section 2.3.

and perhaps this is an unintended one – that the ‘me’ meant to be involved has been predeterminedly separated from the body *prior* to the phenomenological investigation that yielded the three dependent traits.²⁵⁵ What I mean in this is that by using such language Dainton is showing that he has, before actually examining the phenomenology, already decided that what is paramount is only the mind, else there could be no ‘ownership’ type terminology such as the phrase ‘your body’ reveals. If the body is a part of the self then the self does not possess it, rather there is no split of ‘your’ and ‘body’ to be had as they are co-definitional. Is Dainton, possibly inadvertently, arguing from his conclusion? He might not be; it is possible that such instances are simply a matter of how he has related the material, and that when he actually conducted the analyses which he only later wrote about his thinking was clearer than it appears in the final textual description, but regardless of whether or not that was the case we readers are again left with a strong smell of dualism. This is confusing though, for it is hard to see how a stream of consciousness/C-system could have any of the requisite ontological aspects that a properly dualist account would require,²⁵⁶ and anyway Dainton certainly does not mean to be arguing for a dualist self (note again his interchangeable use of ‘subject’ and ‘self’ that is indicative of an anti-realist approach). By championing the mental qua his C-systems Dainton may just wish to stay there, on the mental as all that is necessary (and sufficient) for a self, but as we find these selves so considered within physical bodies getting rid of those bodies proves to be a real challenge, and in my judgment it is a challenge that is only partially met by his arguments, as I will try to show. The charge, I think, or the doubt, of a hidden or unintended dualism seems to follow Dainton’s self-view as he goes on.

²⁵⁵ Dainton, *ibid.*, pp. 206-209.

²⁵⁶ Dainton does argue that C-system subjects qualify as substances due to being experientially unified and interdependently co-conscious (and/or potentially co-conscious), thus yielding a ‘quasi-organic unity’, yet even granting substancehood here it is unclear that such would qualify metaphysically in the way that a dualist self would need to, and it is also beside the point. The main point being, of course, that though Dainton is not making a dualist case for the self an entirely insufficient dualism is nevertheless what seems to result from his arguments; see *ibid.*, p. 347.

In a discussion of primary and secondary possession regarding subjecthood/the self and the body that sustains it, Dainton claims that the body/animal/human animal possesses subjecthood in the secondary sense due to one of its parts – the C-system – possessing it primarily, and he gives an account of the natural process of evolution whereby C-systems arise and then the organisms with those C-systems become in effect ‘possessed’ by them.²⁵⁷ This again is meant to highlight the necessity and full sufficiency of the C-system as against the insignificance of the biological physical form even though said form houses or contains the C-system in question. As a counterexample to this we might entertain the thought experiment whereby an artificially produced and non-organic brain supported C-system were built (computerized, say, or made with silicon rather than carbon atoms), perhaps using the same nanorobots and other equipment employed in Dainton’s envatted brain examples. If such were possible, and given the range of what Dainton is willing to consider possible in his thought experiments it does not seem unfair to put this example forward, would Dainton also hold it to be a full self/subject in the manner in which he wishes to define the self? Given his position on the body I cannot see how he could avoid this conclusion – organic or not the body seems to mean nothing on Dainton’s account – and if he must accept a full self/subject here then all of the same body-based objections that Legrand raised about lived subjectivity and bodily self-consciousness²⁵⁸ that were used to argue against Strawson’s account would apply to Dainton’s as well, in addition to the many bodily and embodied issues that were considered in the previous chapter.²⁵⁹ Dainton’s self is thus an extraordinarily simple one; too simple, one doubts, to really account for the widespread phenomenology of human experience. We will see this drawback more and more as we delve deeper into Dainton’s arguments and examine the

²⁵⁷ *ibid.*; on the primary/secondary issue see pp. 230-231, and on the evolutionary account see pp. 234-235.

²⁵⁸ Legrand, *op. cit.*

²⁵⁹ Chapter 2 Section 2.3 (‘Bodily and embodied issues’).

psychological side to his C-systems. Let us continue.

Just as he is willing to disregard the physical body in his account of the self, Dainton is also ready to radically simplify even the psychological aspects that are usually considered to be necessary within self-views focused on the mind, to the extent that an experiencer is not strictly required. A conscious mental life could, Dainton argues, consist of only a solitary experience: ‘But, from the vantage point of the C-theory, subjects are simply collections of experiential powers, and if simple subjects are possible, then a subject can exist in the form of a nothing more than a capacity for a single simple form of experience.’²⁶⁰ Dainton then expands on this and reduces his account much further than Parfit, or even Strawson, did:

An isolated experiential power capable of producing a solitary sensation is not *just* an experiential power, it is a subject in its own right, a very simple subject. Or at least, this is the case if we accept the C-theory in its broad guise, which equates subject persistence with persisting experiential potential.²⁶¹

This line of reasoning leads to the conclusion that if one’s stream of consciousness is reduced to a single sensation along with one’s capacity for consciousness correspondingly reduced, ‘then one would still be in existence.’²⁶² Note that the ‘one’ in question here is a full human self; this is not an abstracted ‘thin self’ or ‘experiencer’ as Strawson would argue for in his reductionist account, it is as much a human self as the self now reading this. Dainton illustrates what he thinks some of the implications of such a conclusion might be by musing on a VR-4 machine whereby a person’s stream of consciousness enters the

²⁶⁰ Dainton, *op. cit.*, p. 239.

²⁶¹ *ibid.*, pp. 245-246; emphasis in the original.

²⁶² *ibid.*, p. 246.

phenomenal experiences of a dolphin, a bat, a dog: if that person ends up completely phenomenologically ‘being’ a dog – if continuity of the stream of consciousness is preserved – then, Dainton asks, ‘is it not clear that the doggish experiences are one’s own?’

This is an interesting case, and what makes it interesting, I think, is the fact that for the ‘doggish experiences’ to be ‘my (one’s) own’ seems to require that I exit the VR-4 machine as the same (fully human self) ‘I’ who entered it, for otherwise the only stream of consciousness to speak of in this example would be the virtual ‘dog’s’ stream of consciousness. Without a human before and after there is only the machine-based dog. If therefore a return to the human ‘me’ is necessitated to later be able to speak of ‘*my* doggish experiences’ while in the machine then what is the ‘I’ who acquired those ‘doggish experiences’? Did that ‘I’ continue as a fully human self during its ‘doggish’ time? How could that be so? The experiment does not seem to prove that a C-system is enough for a self as much as it merely implies that a C-system could be skillfully manipulated within the confines of a self. But then, what is really being demonstrated here? After all, I can put on a pair of virtual reality goggles and play a video game that gives me a dog’s (or dolphin’s or bat’s or tomb raider’s, et cetera) point of view and then when I have turned the machine off consider myself to have had those experiences – yet whatever my self is it did not actually change during any part of the time I was playing. I did not become phenomenologically anything other than myself playing the game.

It is true that we can give Dainton the benefit of the doubt by agreeing that the VR-4 machine is a much better machine than a pair of virtual reality video game goggles, and that accordingly the experiences garnered are much richer. Nevertheless, if within this richer VR-4 experience there is no human re-emergence from the machine then ‘I’ have long since ceased to exist, having lost all emotional and meta-levels of ‘me’. There would

only be a human brain that now believes itself to be a dog's brain. Would this qualify as a self? Would the experiences undergone still be describable as 'one's own'? Who would then be doing the describing? Surely not the person still strapped into the VR-4 who has lost touch with all of their previous humanity, and for us to speak of that person as now having 'doggish experiences' but still 'being them' amounts in real terms to the equating of their self purely with the physical body stuck inside the machine since all mental aspects that were formerly human have become entirely doggish – and as we have already seen Dainton does not count the body as relevant to the self, thus that cannot be the response he would make here. Played out thusly we can see this as a strong objection to Dainton's conclusions about what is needed for a self.

Approached from yet another angle an entirely different problematic area with this account appears: while having the complete phenomenology of being a dog there would yet need to remain with the subject their extant meta-analytical abilities during the machine-induced experiences of doggish perceptions and sensations, otherwise – if there were a full flip into a dog brain – the time allegedly spent as a dog would just be a blank upon disconnecting from the VR-4 machine and returning to one's normal body. Considering that further point though raises yet another objection, for the physical and the psychological cannot be so easily dissociated, and thus without a dog's actual physical brain structure (the specific way in which a dog's brain is formed and internally interacts) how could the phenomenal experiences that are fed into the human brain connected to the machine be properly interpreted? Would the human subject's emotional and intuitive structures (e.g. such as found in the limbic system, brain stem, et cetera) still be (confusedly) providing feedback and judgments for the doggish perceptions supposedly received? What would it mean to have sensations in the sense of their being 'dog' sensations? Presumably this would entail something along the lines of feeling that one has a paw and that the paw is on the grass or

engaged in running after a cat or the like; but again, how would a human brain respond to any of those stimuli? Certainly not in the same way that a dog's brain would.

Given that a dog's (or a dolphin's or a bat's) brain is physically structured very differently from a human being's, and that the degree to which such psychological aspects as Dainton himself mentions²⁶³ vary widely amongst even evolutionarily closely related species, it seems incredible that such thoughts could even be entertained. A 'doggish me' only seems remotely plausible if we assume that a stream of consciousness is nothing more than awareness itself, and although Dainton does seem to make this assumption doing so leaves out all of the many non-overtly experienced or aware (that is, not actively noticed) elements that feed into our streams of consciousness and thereby – via that feeding into – exert influence. Dainton's argument here can only be said to work because the model it is based on is seriously flawed, seriously deficient. Consciousness, as has been discussed and will be further discussed in the below, is a multi-layered system that includes not only what we are thinking about or actively noticing/experiencing in this instant, but also the preconscious elements such as emotions, intuitions, preferences and outlooks that result from both one's genetics and one's upbringing, as well as the lasting consequences (including again the affective and intuitive results) of the choices one has made. In short, Kristjánsson's Sets 1, 2, and 3 as we have altered them. Let us not forget that a dog would have his/her own versions of these as well, and those versions too would be based both on the dog's brain and its workings and on the dog's personal past. All of these many elements are founded in bodies and in embedded worlds, and although each brain processes these they do not stem solely from the brain, they go into and come out of it. The relationship is necessarily twofold. A machine like the VR-4 would not yield a different set of 'my

²⁶³ Dainton however claims that since these aspects he mentions are non-experiential they are therefore safe to ignore: 'Non-repressed propositional attitudes, many of the more complex emotions, certain character traits, and mental abilities'; *ibid.*, p. 174. As argued I do not think such can or should be disregarded.

experiences', it would yield a different 'me', a different self, as the self-constitutive, self-comparative, and self-conscious aspects respond and shift once provided with an entirely new stream of phenomenal experience and the corresponding psychological (including emotional and intuitional) adjustments are enacted. As the previous chapter sought to establish, the self is always a work in progress.

Dainton does consider a general objection to neo-Lockean accounts that falls along the lines I have made here. The charge analyzed is that such accounts try to separate capacities that define selfhood from the subject of which they are a part (i.e. the mental out of the body), and that such capacities cannot be individuated out of their subjects to try to define the self.²⁶⁴ Dainton responds to this by stating that capacities can be so individuated by reference to a brain and/or brain part and therefore need not turn to the subject.²⁶⁵ In this reply I see a degree of glossing and a failure to discuss the same thing. The objection itself, in the form that Dainton deals with it, seems to equate 'subject' with 'whole person' and hence the charge that the self's definitional capacities cannot be removed from the body in which they are based (or reside or are generated or form a part of, depending on how one wishes to define the self in question). Dainton's response, however, states that if we speak about the brain (or brain part) then we need not speak about the subject, which again appears to indicate that 'subject' is here being used to refer to the 'whole person' (in the 'whole physical person' sense; brain plus all the other bodily parts), yet all along Dainton has used 'subject' interchangeably with 'self' in arguing for his self concept, a self concept that has no use for and is dismissive of the body. Has Dainton's terminological usage shifted or is he simply conceptually imprecise on this point? Either way it does not strengthen his case.

²⁶⁴ The objection, as considered by Dainton, comes from E.J. Lowe, *Subjects of Experience* (Cambridge: Cambridge University Press, 1996).

²⁶⁵ Dainton, *op. cit.*, pp. 349-350.

Moreover, if Dainton's phenomenal self is really only meant to be about the brain (and not mind) then it is still more than just being about a stream of consciousness, for in that case it would have to include the brain's physical elements, the functioning of which contain the preconscious and conscious elements previously referred to. If, however, it is only meant to be about a brain *part* then that part must be what is generative of that stream of consciousness. What part is that? No one knows, yet as we have repeatedly discussed in the foregoing and will discuss again in the next two chapters, consciousness does not seem to be generated by a single part or place of/within the brain. Even so, if by this Dainton means to say that his self is really about that part – if it were to exist – that produces one's stream of consciousness, then it is still again about additional elements beyond just that stream and once more necessitates a physicality of some manner. By failing to draw a clear distinction between the self and the subject – a distinction which matters more here than on Strawson's account as Dainton does not even require an experiencer to be present – Dainton seems to have doomed his self-view to be inadequately prepared to answer charges such as the one he considers, and that deficiency is evident in his response.

On the whole, then, the two contemporary phenomenologically-based accounts of the self that we have considered have, upon analysis, turned out not to be able to fully account for the entire phenomenology that we actually experience in our daily lives. In the next section we will therefore return to consider what phenomenology's founding father, Edmund Husserl, had to say about phenomenology as a means of study and analyses in order to try and resituate ourselves for the remainder. With our methodology thereafter sufficiently refreshed and reinforced, we will turn to consider a thinker who was historically closer (very close) to Husserl and who also had interesting and enlightening insights to provide

on the nature of the self: Martin Heidegger. We will merely²⁶⁶ glance at his self account though, for although it is interesting what will be of more consequential value for our later conclusions will stem from Heidegger's metaphysical notions, and hence the need to acquaint ourselves with his work.

C. Phenomenology 3: Going back to Husserl

Phenomenology was created to be a new philosophical approach and an answer to the perceived crisis that unfolded from the collapse of the 19th century's largely mechanistic and determinedly progressive worldview that the early tragedies of the 20th century gave rise to (and which was dreadfully reinforced by the latter 20th century). As the founder and longtime figurehead of the movement, Edmund Husserl (1859-1938) wrote and lectured on a wide variety of topics, many of which fall outside the scope of the present study. What this section will therefore undertake to do will be simply to provide an overview of phenomenological methods of investigation as Husserl envisioned them and of how such relate to the question of the self, particularly with regards to any potential bearing on the self concept being argued for here. In doing so we may come to understand how those using alternative or adjusted versions of Husserl's methodology have arrived at different conclusions regarding the self, and we will also be better prepared to conduct our very brief survey of Heidegger's thought as he studied under Husserl directly and had much to say regarding Husserl's work and how he wished to – in his view – improve on it. Since phenomenology is immediately concerned with the experience of consciousness this section will additionally, it is hoped, put us in good stead for the more in-depth analysis of how consciousness relates to the self that will form the bulk of the following chapters.

²⁶⁶ I use this term here purposely – and playfully – to pay tribute to Heidegger as he used the word so very often himself.

Husserl created an ontological picture of the world that divided and subdivided objects into types, essences, and regions. Objects were either those of Fact (real in time or spacetime) or Essence (ideal), with the Essence category then further broken down into Formal and Material. The Material essence grouping contains three further subgroups (called ‘regions’): Nature, Consciousness, and Culture/Spirit.²⁶⁷ As David Woodruff Smith writes, pure phenomenology ‘studies experiences while considering only their properties that fall under the essence Consciousness, excluding consideration of their properties that fall under either the essence Nature or the essence Culture.’²⁶⁸ This involves Husserl’s famous ‘bracketing’ (termed *epoché*) of the features of an object and/or experience that define its place in nature or culture, focusing only on the consciousness *of* an object²⁶⁹ and how that object is (ideally) presented to me, the sense it provides and the quality of the experience itself. As Smith explains:

I am to ‘bracket’ the question of the existence of the object I am currently seeing (or desiring or thinking about), and so I am to turn my attention instead to my *consciousness* of that object, to the way I am experiencing it, proceeding to a phenomenological description of this form of experience... Husserl’s phenomenological method ‘brackets’ the object of consciousness – and the surrounding world in general – in order to shift our focus on to the sense or meaning through which the object is experienced.²⁷⁰

This is somewhat like the nuance a word attains when we put it in quotation marks/inverted commas, where the content we are talking about ceases to take precedence

²⁶⁷ On this scheme there is also a third type of Meaning/Sense (ideal content of intentional experience) in addition to Fact and Essence.

²⁶⁸ David Woodruff Smith, *Husserl*, 2nd edn (Abingdon, OX: Routledge, 2013), p. 159. A helpful table of Husserl’s ontological categorization is given on p. 150.

²⁶⁹ I.e. awareness: Husserl’s use of ‘consciousness’ follows the standard mode of implying ‘attention to’ or ‘noticing’; throughout this section the term should be taken in that way.

²⁷⁰ *ibid.*, pp. 189 and 229-230; emphasis in the original.

and instead we turn our attention to the word itself and its usage in the case at hand. It is a shift in attitude from seeing X to *how* one sees X, what it is like (to me) to see X, the experience of one's consciousness of X. Such is of course necessarily relative in certain ways to each person, and to the possibilities open to that person for how they may experience X based on the kind of background factors we have discussed that compositionally inform the person (such as those of present emotional state, personal genetic inheritance, style of upbringing, held intuitive reactions and judgments, parents and siblings, et cetera) and the world in which they are embedded (sociohistorical, geographical, epochal factors, et cetera).²⁷¹ This range of possibilities Husserl very appositely refers to as a 'horizon',²⁷² and we will later return to his thoughts on the matter.

Two other considerations of how one sees X (or one's approach to X) are also of importance here: the first is the transcendence of the thing in question, and the second is the necessary one-sidedness of perception. Regarding the experience of one's approach to a physical thing, Husserl writes that consciousness cannot grasp directly the entire immanent qualities of it, and this is due to consciousness having its own immanent qualities which change from moment to moment affecting how the approach unfolds, thus the fullness of the thing under observation or awareness, in Husserl's words, 'is said to be, in itself, unqualifiedly transcendent. Precisely in that the essentially necessary diversity among modes of being, the most cardinal of them all, becomes manifest: the diversity between *consciousness and reality*.'²⁷³ Put differently, although the physical world in which we exist has its own reality (as opposed to, say, how an idealist theory might have it), we stand in that reality behind a veil of filters or blinders, and such are both unavoidable and of our

²⁷¹ These concerns were covered primarily in Chapter 2 Section 2.5 ('Certain uncertainty, randomness, and limited choice').

²⁷² Ontologically this falls in the region Consciousness under the essence Material and type Essence.

²⁷³ Edmund Husserl, *The Essential Husserl: Basic Writings in Transcendental Phenomenology*, intro. and ed. by Donn Welton (Bloomington, IN: Indiana University Press, 1999), p. 73; emphasis in the original. The quoted section is taken from Husserl's *Ideas I*, Section 42.

own ‘making’ in the sense that they simply come with the territory of being a human animal. This leads to the second consideration just mentioned: perception’s one-sidedness. On that Husserl follows his conclusions on transcendence with some on appearances and embeddedness, stating that:

A physical thing is necessarily given in mere ‘modes of appearance’ in which necessarily a *core of ‘what is actually presented’* is apprehended as being surrounded by a horizon of ‘*co-givenness, which is not givenness proper, and of more or less vague indeterminateness.* And the sense of this indeterminateness is, again, predelineated by the universal essence of this type of perception which we call physical-thing perception... Necessarily there always remains a horizon of determinable indeterminateness, no matter how far we go in our experience, no matter how extensive the continua of actual perceptions of the same thing may be²⁷⁴

We are thus unable to grasp true essences, as it were, but we are able to take hold of our own internal experiences, of our ‘consciousness of’s. Does this mean that we are hopelessly lost, adrift in an unknowable world? It might seem to imply that, or anyway conclusions similar to that, but even if so what phenomenological practice allows one to do, Husserl asserts, is to access pure consciousness in its absolute state: ‘though we have “excluded” [bracketed] the whole world with all physical things, living beings, and humans, ourselves included...we have not lost anything but rather have gained the whole of absolute being which, rightly understood, contains with itself, “constitutes” within itself, all worldly transcendencies.’²⁷⁵ This is meaningful. Having isolated the consciousness of X in one’s investigation from all other considerations one is left only with the

²⁷⁴ *ibid.*, p. 75; emphases in the original; taken from Husserl’s *Ideas I*, Section 44.

²⁷⁵ *ibid.*, p. 84. This portion is from Husserl’s *Ideas I*, Section 50.

‘phenomenological residuum’ (as Husserl phrases it), and that is consciousness itself. This methodology may not be able to teach us much about things like quarks and photons and the nature of the universe as it exists outside of all human considerations (the purportedly ‘objective’ point of view), but it can teach a great deal about our own place in that universe as we live our lives and experience our experiences.

Employing these ideas, what might the implications be for the self? We have presented a self concept that entails its being both substantive but always in a state of becoming, realist yet anything but static. Given that, and taking as our conceptual definition the formula: self = Set 1 + Sets 2 & 3 + C + B_p,²⁷⁶ what might an application of phenomenological practices help us understand about the self? In its precision of individuation Husserlean phenomenological investigation allows us to isolate for a detailed examination any one of the aspects of the self on our view. Through doing so we not only learn about that aspect itself and our experience of that aspect, we also gain important self-knowledge. What are my core traits? What commitments do I hold? How do I consider how I am doing when it comes to being the type of person that I want to become?²⁷⁷ This understanding can then provide the first step on a purposive program of self-alteration, of adjusting those traits, those commitments, of assessing personal progress and of making whatever changes are deemed called for (or not). In practical terms, how might one go about this? We have seen in the previous chapter the central role that intuition plays in decision-making and in the generation of behavior, and we have moreover come to understand that intuitions can be shifted first by obtaining a conscious awareness of what they actually are (including not only inborn and species-specific intuitions such as a natural fear of snakes but also those

²⁷⁶ That is, self-constituting ‘emotions’ (core traits, commitments, ideals, aspirations) + self-comparative and self-conscious (in other words, self-directed) emotions + consciousness + bodily presence.

²⁷⁷ These are of course just some examples of the issues that can be taken from the three emotive sets we have been looking at.

which have been acquired through individual pasts) and then by engaging in determined and repetitive practices aimed at transformation. This is using System 2 thinking to adjust System 1 thinking (as Kahneman would put it), or our rational mind to adjust our emotional mind (as Haidt would put it).²⁷⁸ Each part of the self thus provides a potential point of entry to self reinvention. Moreover, given that the self on this view is realist, an alteration to one of its elements would shift the self in its entirety and thereby too shift the personal identity and whole (embedded) person resting upon the foundational self. Husserl's methodology thus grants a potentially powerful tool by allowing one to mentally step out of one's ongoing life and appraise oneself from a step back, focusing on that which may have seemed unapproachable from the unexamined position. It is a means by which one may self monitor, and in the use of that term here I mean to indicate both the narrower sense of 'the self' and the broader sense of a 'whole person'.

Returning to Husserl proper, what were some of the views he expressed on the self? Our concern in this section is primarily with phenomenology as a method of analysis and so we need not dwell too deeply on Husserl's comments regarding the self, but we can note some areas of interest that may help to inform our study. The first of those is that when it comes to the self the case with Husserl is not always clear and certainly seems to have developed over time, taking a particularly new direction following the so-called 'transcendental turn' that his mature work is said to exhibit. In his *Ideas I*²⁷⁹ Husserl writes on the *cogito*, the *cogitare*, and the *cogitatum*, as well as referring to the Ego. The stance he takes there appears to be anti-realist, largely equating the *cogito* with attention, with where one is mentally directed moment by moment. He states that in the natural position we live

²⁷⁸ See Chapter 2 Section 2.5 and Chapter 1 Section 1.1.C ('Certain uncertainty, randomness, and limited choice', and 'Four accounts of the self: The soft realist position'), respectively, for explanations of the two models referred to.

²⁷⁹ The full title of this 1913 work is *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy – First Book: General Introduction to a Pure Phenomenology*.

‘actively’, and that while so engaged in life, ‘I state the cogito, whether I am directed “reflectively” to the Ego and the cogitare. If I am directed to them, a new cogito is alive, one that, for its part, is not reflected on and thus is not objective for me.’²⁸⁰ Husserl points out that we are not always in this reflective state, not always related to an ego-type ‘I’, and that not every *cogito* has an associated *cogitatum*.

On first encounter, this position appears to indicate that the self for Husserl at this stage in his thinking is not an element of lasting substance since it needs to be reflected upon and attention given to it first and then, when so given, its cognates are realized even as it itself fades from view and becomes ‘not objective for me’. This reading seems too to be reinforced by the indication that directed attention leads to (or causes?) a ‘new’ *cogito* to be enacted (or enlivened), yet not automatically connected to anything. There is much room for nuance here, however, as becoming ‘not objective’ might not indicate that anything disappears, and moreover just because an unreflected upon *cogito* is not in ‘active’ view does not necessarily mean that it ceases to exist altogether. There might still be something there, even the very same *some-thing* there, hovering in the background while one’s mind is directed to the task or interest at hand. If this is what Husserl meant then it would line up much more neatly with a traditional concept of the self that accords it the status of one’s ‘true essence’, and also with Descartes’ own views.²⁸¹

Following this second possible interpretation, we might consider that Husserl’s later ideas on the self actually grew directly out of the seeds of these earlier views. Zahavi, referred to above when considering Strawson’s account, has noted that when Husserl’s whole oeuvre is considered the picture that emerges is one where ‘even the anonymous stream of

²⁸⁰ Husserl, *op. cit.*, pp. 61-62, found in Section 28.

²⁸¹ Therefore perhaps the reason that Husserl chose to use this most Cartesian of terms?

consciousness would be unthinkable without an original ego-pole as the center of action and affection’;²⁸² it does strike the reader that Husserl’s ‘I’ in its latter version indeed appears to be quite holistic. In that incarnation mind, body, and cultural persona are not to be thought of as substances in the Aristotelean sense, rather, Smith writes, ‘they are importantly different aspects of one such “substance” called “I.” In Husserl’s ontology, these aspects are *moments*²⁸³ (dependent parts) of the individual or “substrate” that is me.’²⁸⁴ Each ‘I’ additionally has its Natural, Consciousness, and Cultural/Spiritual properties (the regions of the Material essence), all of which come together in the day-to-day ‘life-world’ (embeddedness); we should remember here too that for Husserl events also have properties of these three categories. The self is therefore ‘that being who experiences the acts within a unified stream of consciousness’, and phenomenologically ‘the intentional content or *sense* “I” encodes that unique mode of awareness’.²⁸⁵ Aspects of both Strawson’s and Dainton’s thought are visible here in Husserl, though the whole picture reveals a much deeper concept of self than either of them argue for, as can be seen in Zahavi’s reference to the ‘ego-pole’ – a ‘being who experiences’ from a stream of consciousness point of view – and from this instructive section found in Smith’s work:

Husserl goes on to distinguish the ‘pure’ or ‘transcendental’ I, pure subject of experience, from ‘I, this man,’ with ‘a psychic life in the world’. The distinction is not between distinct objects called ‘I,’ but between distinct aspects or moments (instancing distinct essences) of the one object, I: my experiencing is one aspect or moment of me, my bodily shape is quite another, and the method of bracketing allows me to focus on my conscious experience as such... Practicing ‘transcendental’ reduction, or bracketing, I reflect on my current

²⁸² Zahavi, *op. cit.*, p. 322.

²⁸³ This is one of Husserl’s technical terms and should not be read with its common meaning.

²⁸⁴ Smith, *op. cit.*, pp. 160-161; emphasis in the original.

²⁸⁵ *ibid.*, p. 221; emphasis in the original.

consciousness. In this reflection, Husserl holds, I have *apodictic* evidence of my current experience ('I think') and of my being ('I am').²⁸⁶

The closing allusion to Descartes notwithstanding, Smith gives us a view of Husserl's self concept that is remarkably compatible with the Kristjánsson-inspired self concept that Chapter 2 argued for, containing as it does core psychological and bodied/embodied factors which are fully interdependent. Moreover, we are said to know such items incontrovertibly, and that is so through reflection, possibly indicating that the reflective *cogito* of *Ideas I* is after all closer to a realist conception than an anti-realist one. What is being highlighted is that where boundaries can be drawn they are not of series-type 'I's but rather of aspects all contained in *an* 'I', not divisions around a separable whole but the parts of the whole: the 'moments' that, being dependent, must rely on a foundation.²⁸⁷ The case is still somewhat vague though, and as yet we do not have much of a full definition of what the 'I' in question, with its 'psychic life in the world', might be in detailed terms. Perhaps in the end these particulars are not all that crucial if we can go from a person functioning in the world into their very specific aspects via the practice of bracketing, but then if such are not needed what are we doing in this study? What is the whole province of self research doing? The question of the self seems to demand the level of scrutiny that it has historically received, the query 'What am I?' too intertwined with the related problems of meaning and purpose for most of us to be able to settle for partial answers or to lightly set it aside. Let us therefore continue along the lines where we have thus far been concerned using the investigative practices involved in phenomenology. Up to this point we have primarily exercised the methodology as it relates to mental aspects of the self, and so we will now

²⁸⁶ *ibid.*, pp. 321-322; emphasis in the original.

²⁸⁷ Strawson tries to circumvent this need for a foundation by having the parts (thin selves/minimal subjects) rely only on each other to achieve a (illusionary) sense of diachronicity; to me this is one of the more intriguing elements of his self concept although as argued I do not think that ultimately it can work.

move to consider the body proper.

As related above regarding Strawson's thin self/minimal subject, Legrand explains how for Husserl consciousness (that is, an awareness) of the world and self-consciousness (i.e. embodiedly self-conscious; this compound should not be read here as indicating a consciousness of the self as an independent entity) are contemporaneous, occurring synonymously.²⁸⁸ The idea of the world as experienced through the body is also picked up on by Smith, who notes that Husserl observes that we always experience objects as located spatially in the context of being centered on our own bodies, but that – given Husserl's ontology and its particular placement of the ideal and separation of Meaning/Sense as the third type of object (with Fact and Essence) – 'the raven itself [currently viewed object] is distinct from its properties and its properties may change over time'.²⁸⁹ This is a metaphysical point that has a long and contentious tradition in philosophy and in the sciences, it is a question of summing (object = aggregate of its pieces) versus emerging (object = something more than its members), and Husserl, I think, favors the latter. This is an area worth exploring, and metaphysical issues regarding how a self may be considered substantively real in the sense in which it has been suggested here will be further considered in the final chapter,²⁹⁰ but for the moment it is perhaps best to table those details lest we become overly involved in what must prove at this point in our study to be something of a distraction. Be that as it may, the emergent stance just highlighted can still be instructive in how we might view the interconnectedness of the body and the self, particularly if we wish to maintain the argument that has heretofore been presented. If so, and I can think of no reason not to entertain the idea, then the raven (or whatever) is more

²⁸⁸ Legrand, *op. cit.*

²⁸⁹ Smith, *op. cit.*, p. 212.

²⁹⁰ Emergence as it relates specifically to questions regarding the nature of consciousness will also be considered in the next two chapters.

than the total of its properties, and even should they change (height, weight, location, action, et cetera) the raven would not cease to be categorically (on Husserl's categorical scheme) nor stop being existent as such, as 'the raven'. Applying this to the self on the account that we are arguing for, we could claim that it is an emergent entity in the same or a similar way, and thus it continues despite any number of internal changes that it might undergo. Is this convincing? That remains to be seen, but the possibility is there and with it the impetus to continue our examination.

The very common failure to account for the body in a consideration of the self that we have repeatedly noticed in the foregoing may be related to a criticism that Husserl held for the scientific worldview in general, that the descriptions of objects and nature that the hard sciences such as physics yields 'mathematizes' our experiences of the spatiotemporal and abstracts away from the roles our bodies play in perceiving, say, the movement of things in nature. This tendency also abstracts away from those same objects' necessary intersubjectivity when considered from the point of view of a perceiver/experiencer whose every perception and experience comes via a body. Is a truly objective view possible from such a starting point? That is a poignant question and one worthy of much discussion,²⁹¹ and though such too would be outside the scope of this section (it will be returned to in Chapter 6), merely raising it reminds us that when we investigate the self we ought to remain ever cognizant of being in the position of standing inside and looking further in, never outside and never without preconceptions. Phenomenology as a methodology is very helpful here for it is a tool (or a set of tools) that allows one to maintain an embodied grounding even while focusing on the consciousness of X as one turns from within to

²⁹¹ For an engaging look at this topic see Paul Feyerabend, *Against Method*, new edn, intro. by Ian Hacking (London: Verso, 2010). There are some remarkable antecedents to be found to this whole line of thought in Nietzsche's *Beyond Good and Evil* (and indeed within his work taken generally), as well as to the intuitive judgment model as presented by Haidt, Kahneman, Greene, and others; Friedrich Nietzsche, *Beyond Good and Evil: Prelude to a Philosophy of the Future*, trans. and comm. by Walter Kaufmann (New York: Random House Inc., 1966).

deeper within and analyzes experiences.

Many of the varying theoretical arguments on the self, both in the realist and anti-realist veins, could be said to be subject to this same ‘mathematizing’ tendency, to the preconscious prejudices (intuitive judgments) that modern educations have bequeathed and which result in a thinking done solely in the terms and categorizations of scientific training. This has a necessarily narrowing effect. Yet in indicating it I do not mean to denigrate the scientific method nor to downplay its importance in our world and in our lives, I only wish to emphasize the unreflective manner in which we retreat to it even when considering problems – like that of the nature of the self – that require a resistance to this mathematization, for such cannot be considered in the isolated way that the sciences generally demand. It is right and good to set up an experiment on particle physics that seeks to create the conditions present in a vacuum, yet the self quite obviously will never exist in a vacuum, particularly as long as we wish to focus on human selves. Perhaps Strawson or Dainton are correct in asserting that there could be a creature which we might call a self or which we might say has a self that is only a single moment of experience (with or without an experiencer) – the universe is a wondrous place –, but that would not and does not account for anything like human experience, and so it is perhaps unsurprising that we found both self concepts, and others like them which employ similar strategies, to fail to adequately encompass the human case. Husserl saw this clearly a century ago, but we appear not to have learned the lesson.

Yet despite the importance of the body in human experience and in properly engaging in phenomenological investigations, phenomenology as a method does appear to remain

almost entirely mental in its focus on consciousness (as awareness²⁹²) and in its application of the ‘bracketing’ method, wherein one reflects on the consciousness of one’s experiences and not on the physicality of them (should such be included) as sensually felt. The line is perhaps hard to draw: how could one feel the touch of the wind without being conscious of that feel? To *feel* it with the senses means to engage the mind in recognizing that feel, else it would remain as preconscious data and as an unrecognized potential influence. Admitting this enables learning from it, for thinking of this sort does help shed light on the way preconscious inputs might affect us. Moreover, when making use of phenomenological analyses we do not forget or ignore the world that we are embedded in; rather we turn our attention more specifically to ‘the content through which we experience things in the world.’²⁹³ We go within, and to a specific place within, to that which we focus on to the exclusion of all else in our bracketed inquiry.

Still, this is not a purely internal and private mode of study, for phenomenology calls on us to, and takes its force from, describing the constituents of the conscious experiences that we have turned towards, thus producing statements like <I see the clock tower of a high school about five hundred meters away>. In the forming or reading of a statement such as the preceding, Husserl argues, I understand how it would be to myself have such an experience. At this point however a doubt naturally comes to mind, for such a simple declarative seems to be missing what consciousness could really be said to be. If I have had no similar experiences (or capabilities), then how useful is a description like the preceding? Perhaps I have never seen a clock tower, and although I do know what ‘clock’ and ‘tower’ individually mean I have trouble combining them pictorially. Or perhaps I do

²⁹² The reader will recall that on our model where consciousness is considered to always be present to at least some degree (whether with awareness or not) the term has broader implications than Husserl’s usage of it indicates. Again, in our examination of Husserl’s methodology ‘consciousness’ is being used in the sense in which he employed the term – essentially as equivalent with ‘awareness’ – and not in the way we understand it. This topic will be taken up again in the next two chapters.

²⁹³ Smith, *ibid.*, p. 240.

not know what a ‘high’ school is and my only references for such places as schools are in terms of primary, secondary, and tertiary. Or perhaps I cannot think in metric terms and so have no idea what ‘five hundred meters away’ might mean; is that near or far? How near or far? The ‘away’ seems to indicate distance of some kind, but to what degree? Or perhaps I have never seen anything at all and simply cannot relate to the entire sentence. In any of these cases or others, how much of the roots of what conscious experience entails for the manner of creatures that we are can really be said to have been captured? ‘I see that’ or ‘I think that’ cannot be held to fully describe what it is to see or to think in the absence of such previously held commonalities, including – importantly – the linguistic conventions and shared meanings and associations involved. The question remains: Just what is *conscious of*?

The answer to our doubt may lie in the fact that we have (once more?) mistakenly looked in the wrong place, from the wrong angle. Smith writes:

phenomenology studies individual experiences only in order to develop laws about consciousness in general. The phenomenologist’s interest is not in my experience just now as I see ‘that eucalyptus tree across the street,’ but rather in the structure of consciousness typical of visual perception – and more general forms of experience such as intentionality.²⁹⁴

Phenomenology itself need not really even say what conscious (including conscious of) is as its concern is with our human experiences *of* consciousness of. This is not a definitional pursuit; it is a descriptive one and will therefore, as descriptive, depend in some measure on what is shared between you and me. The reader will notice how important joint or communal meaning is here, for without it any attempt at any kind of description will fail.

²⁹⁴ *ibid.*, p. 244.

In Husserl's ontology meaning is ideal – in the philosophical sense of the term –, and as such it functions as the medium by which the intentionality of one's acts (broadly defined, including thoughts) is directed towards objects (also to be understood broadly, including for example others). Smith again summarizes the argument this way:

experientially, phenomenologically, our consciousness is propagated through meaning toward the object, so that I am visually conscious of that tree across the street, without being in any way aware of the meaning through which my consciousness is so directed... we are conscious of things – we know or 'intend' things – only *through* structures of sense that present or prescribe those things in particular ways. Yet we are not aware of the sense through which we experience an object until we step back from the experience and abstract its content.²⁹⁵

Thus we might say that for Husserl consciousness involves more than just the functioning of my human form in the world, it also involves meaning in the particular sense of conceptualizing received data through willed awareness of it or willed attention to it: by that process something is known. In other words, the data by which and in which we live comes in the bifold form of ideal-material, and we come to be conscious of (that is, learn about, know about) that data when we purposely apply the ideal to the material, when we 'step back from the experience and abstract its content' as Smith puts it.

In taking such thoughts to the problem of the self as we have so far considered it, an interesting query presents itself: Could the self be understood as an ideal medium of our experience as it is experienced? Could the self be a 'real' substance that undergirds our ongoing ideal-material experiences without having materiality itself? Could it be the

²⁹⁵ *ibid.*, pp. 246 and 249; emphasis in the original.

‘meaning’ medium in which we live, itself also flying under the radar of awareness until we turn our attention to it? The question seems very much to hinge on what we are willing to consider ‘real’, and in what way we might wish to define or conceptualize ‘reality’. This is an extremely important aspect of this study but its analysis must be saved for the final chapter. However, at this stage in our discussion it is still illuminating to realize that on the self-view being argued for one could step back and abstract (from an embodied/embedded point of view) the content of one’s self, one could reflect on – and indeed emotionally one is already doing this – the constitutive elements of one’s self, and such is possible consciously (e.g. harnessing awareness through one’s ever-present consciousness) precisely because consciousness is a constitutive element of the self. This process is available to us due to the layered nature of consciousness as possessed by human beings, due to our innate ability to abstract, to employ our System 2/rational tools and not simply remain bound to our System 1/automatic functioning.²⁹⁶

It is at this juncture, finally, that we come back to Husserl’s significant concept of ‘horizon’. One’s ‘horizon’ is the range of possible understandings that are open to one, and which differ based on the particular past one has; this implies of course the embedded influences that we have been discussing. Each person’s ‘horizon’ is thus constrained in important ways by that person’s own previous experiences and, as Smith writes, the ‘content of my implicit – often vague and indefinite – *background* ideas about such objects, including beliefs, expectations, and practices.’²⁹⁷ We see again the magnitude of intuitive and affective influences, and though such are not pointed to explicitly the psychological research²⁹⁸ covered in earlier chapters finds here philosophical reinforcement – presciently

²⁹⁶ As Chapter 5 will argue, consciousness is not just this abstracting potential, and there are degrees or levels of consciousness such that we might apply the term ‘conscious’ in ways and to creatures for whom abstraction remains a biological improbability or (perhaps) impossibility.

²⁹⁷ Smith, *ibid.*, p. 274; emphasis in the original.

²⁹⁸ See especially Haidt (2001), *op. cit.*, Haidt (2012), *op. cit.*, and Greene, *op. cit.*

so, we might add.²⁹⁹ If we apply this notion to the present study's self concept, we see that the self as an ongoing and ever-changing project – always being built, shifted, transformed, even when not reflected on – too could be described as having its own 'horizon', and that the many environmental and situational factors that feed into one's self could not, on such a consideration, ever really be separated out, just as bodily issues could not be thought of in the distinct (or discounted) way that a purely mental account of the self might do.

Moreover, Husserl's 'horizon' also includes the epistemic grounding that one has, a grounding that is itself embedded in the knowledge held by the specific 'world' around one, a knowledge that, rooted in language and culture as it is, has been handed down from one's predecessors and which thereby has attained layers of 'sedimentation'. All of what one knows and even the manner in which one knows consists of and sustains this further background, one that 'takes the form of my implicit sense of the life-world'.³⁰⁰ Would our self-view also contain such epistemic considerations? It seems highly likely – almost inevitable – that it would, and certainly the idea that received language and culture would exert their unseen hands on how we know could fit into the three emotional sets that we have been examining and that form a core part of our definitional self.³⁰¹

By way of review and to close out our discussion in this section, on our account the self, personal identity, and the whole person are as interdependent upon each other as the constitutive elements of the self are for the self: to the root self level we added the

²⁹⁹ Heidegger too strongly emphasizes the import of such elements, as the below will discuss.

³⁰⁰ The term 'life-world' indicates one's surrounding world considered in its totality – full embeddedness, as it were; Smith, *op. cit.*, p. 332.

³⁰¹ As a reminder: Set 1 consists of self-constituting 'emotions' (preferences, outlooks, genetic inheritance, upbringing, considered choices such as goals, aspirations, et cetera), and Sets 2 and 3 are self-comparative and self-conscious emotions (the affective and intuitive reactions that monitor and respond to the whole, and maintain, shift, sustain, and/or adjust that whole).

contingent facts about the body,³⁰² other contingent facts, and social feedback to arrive at one's personal identity, and to that we further added embeddedness to arrive at the whole person. Yet when we consider ideas like Husserl's 'horizon' we can see once more³⁰³ that embedded factors must inform the self just as elements of personal identity also inform the self; the entire relationship loops. That does not mean however that the self – as that foundational core – cannot be investigated, just that it cannot be isolated and pulled out from its world and the influences it thereby receives. It probably indicates too the inadvisability of an approach that attempts to seek out the self as something static and antecedently existing and which must be found or discovered in some way. Rather, on our view, the self (and personal identity, whole person) ought to be thought of as an ever emerging and wholly dynamic phenomenon, and if so then phenomenological methodology, with its emphasis on the consciousness (awareness) of experience, appears to be a fitting means of continually studying one's self throughout one's life.

D. Phenomenology 4: Heidegger on the self

Although Martin Heidegger did not write very much directly on the self as such, his lifelong work on being, and particularly his core effort *Being and Time*, touched centrally on issues of the self. His concept of *Dasein* has moreover been much discussed in contexts concerning the self specifically and/or in regards to the role(s) of the whole person and the interplay between a person, being, world, society, et cetera.³⁰⁴ As might be expected,

³⁰² On our definition the self itself requires only bodily presence, and so such contingent facts go beyond that mere presence.

³⁰³ As we did in Chapter 1 Section 1.1.D ('Four accounts of the self: The contextualized soft realist position') and Chapter 2 Sections 2.2.D, 2.3, and 2.5 ('Psychological issues and the self: Selves, personal identity, and whole persons', 'Bodily and embodied issues', and 'Certain uncertainty, randomness, and limited choice', respectively).

³⁰⁴ Hereafter the term 'Dasein' will not be italicized as it has at least entered the philosophical lexicon of the Anglophone world even if it has not entered into general English usage. The term itself may not be precisely translatable into English but it can be read as standing for the self, although more properly from a linguistic point of view it is perhaps better rendered as 'existence'. Dasein is characterized by selfhood and Heidegger refers to such as 'evermineness', which for him comes in the three modes of authenticity, inauthenticity, and averageness/indifference; see Richard Sembera, *Rephrasing Heidegger: A Companion to Being and Time*

Heidegger's work is fascinating, complex, and difficult, and so in order to approach it we will need to begin with some preparatory comments on Heidegger's thoughts concerning the place of the individual in and related to the world. What this section will attempt to do is not to provide a definitive analysis of Dasein, but rather simply to give the reader enough of a foundation for our later metaphysical explorations on the reality of the self – traveling as we will along routes that take much from Heidegger – to be approachable and, ideally, philosophically sound.

The notion of comportment is an important one for Heidegger, and in his usage it indicates the possible ways that one has of being towards the things (widely understood) in one's environments. Comportment involves how one is or may be pertaining to everything around one, and ontologically is found at a level much deeper than that of perceptions as perceptions on Heidegger's view necessarily include theoretical superimpositions via later re-interpretations, whereas comportment precedes such. To use an illustrative analogy, if perceptions (as here) are the music one makes in the world then comportment is how the strings of one's self are tuned before any notes are struck.³⁰⁵ One is always in some kind of comportment, some kind of being-towards, for one is always already in the world and can never escape that embedded condition. The world and the self are co-given phenomenologically, and for this reason Heidegger rejected any notion of the self as a detached core or an epistemological subject; for him knowing or learning about the world is knowing or learning about the self, and vice versa.³⁰⁶ In an extended exploration of Heidegger's ideas on tool-being, Graham Harman extrapolates this as Dasein understanding from a position of pre-thematic referencing, from, 'the occluded

(Ottawa: University of Ottawa Press, 2007).

³⁰⁵ Here it must not be missed that considering perceptions in this way is in line with Haidt's work on intuitive judgments that was described and discussed at length in Chapter 2; on the centrality and pre-analytical place of comportment for Heidegger see *ibid.*

³⁰⁶ Jesús Adrián Escudero, 'Heidegger on Selfhood', *American International Journal of Contemporary Research*, 4:2 (2014), 6-17.

underground zone of execution’ which means, on Harman’s analysis, that, ‘the second, more widely recognized sense of Dasein (Dasein as the special entity which has an understanding of being) is only intelligible in light of the first (Dasein as actual existence in the world prior to any presence-at-hand).’³⁰⁷

Heidegger stated that this relationship of self and world is necessarily circular, that since one can never be towards (comportment) things outside of the system of references from which one comes to them, there exists a default ‘pre-structuring’ to one’s understanding which can never be avoided, nor can it even be paused.³⁰⁸ This is essentially akin to accepting Husserl’s notion of horizon but taking it much deeper and indeed even throwing out the possibility of a Husserlean bracketing (*epoché*); Harman writes that the ‘real action in Heidegger studies’ consists nowhere else than in this tension, this strife, between an entity’s ‘concealed execution and its luminous surface.’³⁰⁹ Heidegger insists that this is simply how it is for us as the creatures that we are, and his critique of Husserl is found here as well: embeddedness runs too deep for the simple refraining from action on or about a thing to free it enough structurally for one to learn about its being (in other words, bracketing must always prove *insufficient*). The horizon of which Husserl wrote is for Heidegger not only limitless but bottomless, and ours is not a universe of phenomena or substances it is one of references and functions; Harman again: ‘In other words, Husserl is interested in the things themselves *as potential targets of consciousness* [awareness], while Heidegger is interested in the things themselves *for themselves*.’³¹⁰ This is where Heidegger’s ‘world’ breaks away from Husserl’s ‘horizon’, and it is where the student breaks from the teacher. For Heidegger Dasein is inextricably connected with that in which

³⁰⁷ Graham Harman, *Tool-Being: Heidegger and the Metaphysics of Objects* (Peru, IL: Open Court, 2002), p. 42.

³⁰⁸ Sembera, *op. cit.*

³⁰⁹ Harman, *op. cit.*, pp. 127-128.

³¹⁰ *ibid.*, p. 141; emphases in the original.

it finds itself, in which it exists, moves, has its being. This implies further, for Heidegger, that the self must ultimately be viewed in such a way that we would be forced to label it as non-realist on our terms, for Dasein is so bound in and defined by the specificities of the ever-ongoing projects that it is caught up in at any given moment (and these are not necessarily human nor even animatedly-based ‘projects’) that it cannot be separated or even segregated enough to allow for any kind of Dasein-substance to be discerned.³¹¹ There is no spare room here for a self that might remain in place in the face of external change – not even a micrometer’s worth – and all is in a state of constant transition. The world, the universe, the entirety of the cosmos is itself a single substance, a single singedom, a unitarity: it is Being, and Dasein is ‘merely’ there amongst that.

These thoughts tie directly into Heidegger’s special usage of ‘world’ and the almost unstateable importance of embeddedness. For Heidegger the term ‘world’ took on a far more nuanced and laden meaning than it is typically given, and the foundational ramifications of this notion will continue to be seen throughout this section and in the following chapters. Recall that for Heidegger each of us exists within a framework of references and never – *can* never – exist in a manner that is wholly objective or without preconceived concepts and interpretations; the ‘world’ one lives in is therefore not that of the standard subject-object distinction. In fact, there is no pure ‘object’. To Heidegger, the meaningfulness of all that we encounter is contextual and contextualized, everything, including our selves, is part of an interrelated system whose totality is ‘world’.³¹² It is interesting to consider here how this personal point of view – one which is largely unrecognized by us as we go about our lives – does not so much give rise to ‘the’ world as it does to ‘a’ world. If we follow Heidegger’s considerations and give embeddedness the

³¹¹ *ibid.*

³¹² Sembera, *op. cit.*; cf. Caputo, *op. cit.*, especially pp. 200-219, and Tiger C. Roholt, *Key Terms in Philosophy of Art* (London: Bloomsbury Academic, 2013), pp. 107-109.

centrality that it demands by this analysis, then to be in ‘the world’ is to be in ‘a world’, to be in *your* world, however much particular elements of that world are shared with others in one’s social milieu (perhaps culturally or subculturally). You or I (Dasein) will always understand what is possible for us from this slightly, or very slightly, different and unique perspective, regardless of how unreflective that understanding might be. Our moods too will naturally further individuate this world for us, and in thinking on that we may once more bring to mind Chapter 2 of the present study and the psychological issues related to the self discussed there. We might also recognize again that the social side to our world will further feed back into our personally held understanding of it – and Heidegger did not miss this social aspect, writing that the day-to-day world is one that is shared, that this is a conviction that we all have, that we start from there and as such it is a ‘brute fact’ of our lives and lived experiences and requires no theoretical explanation.³¹³ That is to say, Dasein can never free itself from its linkages with others, and Dasein, Heidegger tells us, understands this. To be in the world is to be embedded: the self is necessarily and definitionally – compositionally, indistinguishably – thus situated relationally between world and (it)self, Dasein and being-in-the-world.³¹⁴

One final aspect of Dasein’s being, and of being-in-the-world, that needs to be mentioned is that of its future orientation. On Heidegger’s account the self is determined through its decisions and actions, its engagements and interpretations which are projected towards the future, rather than by through a union of the past and present.³¹⁵ Dasein is thus comported as ‘being-self-ahead’, understanding itself in terms of its still open possibilities. In exploring this idea Heidegger adjusts his earlier existential foundation of concern to one of temporality or ‘timeliness’: Dasein is essentially incomplete as long as it yet exists and

³¹³ Sembera, *ibid.*

³¹⁴ *ibid.*

³¹⁵ Escudero, *op. cit.*

therefore remains with (whatever) availabilities are before it,³¹⁶ meaning that Dasein is always a becoming, always a work in progress. This too was found and stressed in our own analysis of the self in Chapter 2 and it will become central to the discussion to take place in Chapter 6, where on our account as well time and how the self relates to time will be considered at length. The focus there will also be on the self as a never-ending and internally-bound project.

Heidegger's thoughts on the self, on Dasein and on being, open the way to some interesting side points which we may make by way of closing out this section. One pertinent aspect that Richard Sembera raises is that philosophy is not and cannot be purely theoretical, that the analysis of Dasein is primarily that of one's own Dasein; if one is not willing to call one's own existence into question then any phenomenological investigation of being will not be possible.³¹⁷ Although *Being and Time* situates itself within the framework of the meaning of Being in general, it is the analysis of human existence that is its pursuit, and the book aims to avoid reverting (or retreating) to the types of general categories that the metaphysical tradition is thought to favor; that is, to resist interpreting Dasein in universalist terms aimed at other entities, or at the human plus other entities.³¹⁸ These thoughts perhaps spell out clearest what is dissatisfying with approaches like those employed by Parfit, Strawson, and Dainton in their attempts to explain away the self via thought experiments that involve human, extraterrestrial, and wholly imagined creatures. By trying to say too much, to conclude too widely and reach too far, their methodologies may turn out to do too little.

³¹⁶ Sembera, *op. cit.*

³¹⁷ *ibid.*

³¹⁸ Martin Heidegger, *Basic Writings: Key Selections from Being and Time to The Task of Thinking*, ed. by David Farrell Krell and for. by Taylor Carman (New York: HarperCollins, 2008); these comments are taken from the chapter 'General Introduction: The Question of Being' by David Farrell Krell, pp. 1-35.

Another aspect that Heidegger teaches, and that also involves the kind of thinking about categorization that we find in Kantian traditions, is that entities ‘in themselves’ (outside of and beyond the world) are undiscoverable by us and therefore incomprehensible to us. We are necessarily connected to the world and the world is a structure of Dasein. What we can discover about entities – but not about their existence as such – depends on us.³¹⁹ This calls to remembrance the species-based limits of imagination that Nagel reminds us we have,³²⁰ restraints that we are wont to ignore yet that nevertheless inescapably characterize the manner of creature we are. Being in itself might be essentially indefinable by us, entities in themselves might be unknowable to us, and given the roughly five thousand year quest to find or define the self (which remains ongoing), we might conclude that it too is beyond our ken. In all three areas, however, we can still make approaches and even progress towards an ever better understanding of what Being, entities, the self, are and mean for us *as us*. In his essay ‘The Way to Language’³²¹ Heidegger seeks, ‘to bring language as language to language’, and in doing so not to lose language as itself into a broader picture of a universal of which it is an instance; this strikes me as a beautiful parallel for self studies and a worthy caution to all of us who wish to come to better terms with the self as itself – and that purely in the human case.

Heidegger warns of the far too common scholarly tendency towards metaphysical categorizing, to only and preemptively seeing everything from the perspective which we inhabit, and to determining all ‘with regard to an already established interpretation of nature, history, world, and the ground of the world, that is, of beings as a whole.’³²² The modern approach to the self, I think, is in danger of falling into the same or a similar trap

³¹⁹ Sembera, *op. cit.*

³²⁰ *cf.* Nagel 1979 and 2012, *op. cit.*

³²¹ Found in Heidegger 2008, *op. cit.*, pp. 397-426.

³²² *ibid.*, taken from the essay ‘Letter on Humanism’, pp. 217-265 (p. 225).

of either being stuck in metaphysics or stuck in a ratio-scientific viewpoint that is entirely founded on a perfunctory materialism. I admit that my own thinking has at times exhibited such trends, and Heidegger's words are ones that I take to heart. I would argue that the results of empirical studies on how we function as biological entities can and should inform our philosophy of the self (most especially from the fields of psychology and cognitive science) as such are focused on the human and how we are as humans, yet if one operates from within that stance there remains the temptation to reductivism, to atomize and tear away layer after layer in an attempt to achieve an assumed core, a center that may be far more elusive than we can imagine or even one that may not be possible in the way we imagine.

Prior to Descartes the self was largely just called 'soul' and accorded some degree of respectful mystery, after him the fashion in vogue seems to be to approach the self philosophically the way a chemist does their work. The present study is guilty – if guilty is the right word – of that too, and while such chemist's tools do exhibit merit it can also be easy by them to lose track of what we are actually after. In his 'Letter on Humanism' Heidegger warns us to look out for that something missing, that which gets passed over: 'The fact that physiology and physiological chemistry can scientifically investigate man as an organism is no proof that in this "organic" thing, that is, in the body scientifically explained, the essence of man consists.'³²³ Nor in searching for that essence can we simply add soul or mind or even character for 'In each instance essence is passed over, and passed over on the basis of the same metaphysical projection.'³²⁴ I do not here mean to claim that studying the self is equivalent to studying the essence of the human, but the applicability of the quoted thoughts should be clear: We must drop our preconceptions,

³²³ *ibid.*, p. 228.

³²⁴ *ibid.*, p. 229.

even those rooted in a scientific methodology which is and has been exceedingly useful, because not only may we miss something important by them, but we may also remain lost in a focus on the past or present, neglecting the future. We may be trying so hard to say what the self *is* that we miss what it *can be*.

3.4 Concluding introduction

This chapter began with a look at a very general possible objection to the self concept that was offered in the previous chapter following some considerations of cognitive, emotional, intuitive, and bodily issues. Two contemporary accounts of the self that make use of phenomenological investigative techniques were then examined and found to be insufficient for reasons similar to those regarding the general representative accounts outlined in Chapter 1. We thereafter returned to phenomenology's founder, Edmund Husserl, to resituate ourselves methodologically and to reacquaint with his thoughts on the self, which had already been covered in part by some of the concerns that Chapter 2 dealt with. Following Husserl our picture was filled out by Martin Heidegger's concept of Dasein and his concern with Being and being. Throughout all of that what proved to be central was consciousness: the experience and the feel and the use and the meaning of which remains a giant question mark. Consciousness calls for further attention.

Our attempt at an equation meant to define the self was: $\text{Set 1} + \text{Sets 2 \& 3} + \text{C} + \text{B}_p = \text{the self}$, where Set 1 consists of self-constituting 'emotions' (such personal characteristics as preferences, traits, ambitions, genetic inheritance, upbringing, considered choices, et cetera), Sets 2 and 3 are the self-comparative and self-conscious emotions (affective and intuitive reactions to the whole self that act to maintain or adjust it), C is consciousness, and B_p entails bodily presence (simply the having of a body, with the deeper details of that applying to personal identity, something we distinguished apart from the self proper). In

our account too consciousness is a fundamental part of the self and yet, beyond some brief words on how it may be structured in the brain and how easy it is to misspeak and mis-think about, we can hardly be said to have given the topic enough care. The following two chapters will therefore attempt to rectify this situation by analyzing more deeply, more widely, and more practically what consciousness may be and what it may encompass for the self in the case of human beings. This will necessarily involve some ontological difficulties and theoretical abstractions, yet an effort will be made to follow phenomenology in its concern for the experienced as it is experienced.

I have argued that most theories of the self have attempted *to reduce to it* or *to discover it* as something antecedently existing, and in some ways at some times the present study has made use of those approaches as well. What we found, however, is that the self is neither a static core (as realist accounts have had it) nor an illusory absence (as the non-realist accounts have had it). The self has been argued to be real but never as just ‘this’, and nor ought it to be considered wholly in the absence of the elements of personal identity and issues of whole person embeddedness that it informs and that inform it in a ceaselessly dynamic relationship. Rather, the self is an ever-changing inner comprehension, an understanding arrived at when we look within and ask ‘Who am I?’, and then apply what we find to how and where we are in the world we are. The answer arrived at may vary from day to day or from what another might say of us, but there can be no one ‘right’ answer, both – or a multitude – are all ‘right’ in the important aspect of the understanding they yield to the holder and bearer of the description given. As Husserl has shown us, we can look at an object from any of a thousand different angles and receive a thousand differing experiences of it, all of which point to something about the same object. Caputo, amongst others, has ascribed such ideas to truth – what is (the) ‘true’ – itself. If I am particularly deluded about myself another’s description of me might be more apt, but that

is a case of better/worse, not right/wrong. How I understand 'me' and how you understand 'me' are not after all empirically-driven proofs and need not be treated as such. To say that the self is Set 1 + Sets 2 & 3 + C + B_p is not meant to be the final word on the subject and even that definition, as clear as it might seem, has wide open spaces for a great variety, particularly given that the compositional emotional sets are in constant motion throughout a person's life, and that consciousness and the experience of consciousness similarly never seems to sit still even for a moment.

I think really that we simply cannot approach the question of the self as if an unequivocal answer were possible, as if we could say with one hundred percent certainty that 'I am X' as Descartes tried with such brilliance to do. That type of thinking, the modernist, rationalist, absolutist type, has in my view stayed with us for far too long as we try to apply what works very well for the sciences to every nook and cranny of our human existence. This is to our detriment. I would suggest that we make ourselves free enough to allow for fuzzy pictures and ill-defined boundaries, that we permit the 'you' and the 'me' to bleed into each other at times. This is not the type of answer we are used to wanting, used to finding satisfactory, but it seems entirely appropriate to the task at hand. In fact, it seems the best we can do, and that phrase – 'the best we can do' – should be read in its most positive sense. In that spirit I give the following forewarning: As we attempt an exploration of consciousness in the next chapters our comprehension of it will likely remain dim and incomplete – there is simply too much that remains unknown –, and yet we may hope that nevertheless some deeper understanding of the self may be acquired, and from there perhaps too some fuller appreciation of the lives we have as the creatures we are. We shall have to try and see.

Chapter 4: Galen Strawson's Panpsychism, Subjecthood, and the Self

4.1 A starting point

To recap, we have argued that the self, situated foundationally and definitionally below personal identity and whole person considerations, should be thought of by the equation $\text{Set 1} + \text{Sets 2 \& 3} + \text{C} + \text{B}_p = \text{the self}$.³²⁵ Each of these self-composing areas has been discussed at length, yet there remains a dark cloud over one of them, an unavoidable fog as the element itself remains shrouded in mystery: that element of course being consciousness. The question of consciousness is a question that philosophers, particularly in the analytic tradition, have almost obsessed over for at least the past century, and if we equate mind with consciousness then we can extend that particularly energetic investigation back many more centuries indeed. I do not and would not claim to have any answers to the questions being asked by the eminent thinkers who are currently engaged in this area, but the present chapter and the next will nevertheless attempt to focus on issues related to consciousness since it is a basic component of the self on our definition, and thus to only give it the attention that it has heretofore received would be amiss.

The following will therefore first consider the question of consciousness' place via a discussion of the currently prominent topic of panpsychism (with its implications for subjecthood), and in the next chapter we will move to a deeper look at what insights modern neuroscience may offer before then taking some of the phenomenological lessons

³²⁵ For the sake of ease of reference, here again is how the equation plays out: Set 1 is made up of one's self-constituting 'emotions' (preferences, outlooks, genetic inheritance, upbringing, considered choices such as goals, aspirations, et cetera), Sets 2 and 3 are one's self-comparative and self-conscious emotions (affective and intuitive reactions that monitor and respond to the whole of which they are parts and that are continuously maintaining, shifting, sustaining, and/or adjusting said whole), C is consciousness, and B_p is bodily presence (or merely the having of physicality as opposed to being specifically physical – that is, being physical in certain ways (i.e. as pertains to contingent facts about the body; such details are found at the higher level of personal identity)).

of the previous chapter into a consideration of issues related to the somewhat (on some accounts) controversial idea of qualia. Our meditation on consciousness will finally conclude by bringing the three covered areas together into an overall analysis of consciousness, qualia, and the self. I hope to thereby provide the required grounding for the first half of our study's final chapter wherein the metaphysics of the self will be considered, with attention paid specifically to how the self might be thought of as a real and substantive entity yet one that is so in an immaterial, and perhaps counterintuitive, way. The goal of these two chapters and the opening of the final one will be to give a summative, fundamental picture of the self before we complete our work with an analysis of the lived self. That is, with the self as in/with time, the self as a work-in-progress. To step out on this path before us we will now examine a case for, and some criticisms of, panpsychism.

4.2 Panpsychism and subjecthood

A. Strawson's total physicalism

The argument for panpsychism that we will consider is Galen Strawson's, and we will focus on his presentation because it is directly connected with his previously examined theory of the self. In what follows the reader, I think, will be able to see quite clearly how in Strawson's early claims for panpsychism the seeds lie out of which his self theory grew. The below will therefore act to shed yet more light on concerns related to the self, especially regarding experience and the self, and thereby help to further fill out our own central conception of the self as an ever-shifting, yet permanent, aspect of each of us.

Strawson begins his case by stating that experiential phenomena are physical phenomena, that they are 'real concrete phenomena' and hence equated with and as physical

phenomena, and that his view is based on our experience of experience.³²⁶ In his introducing of it Strawson points out too that this is not a new idea, citing Hume, Locke, and Priestly as allies, yet we may interject here that as far as this point of pedigree goes – and regarding the term ‘physical’ – some initial caution on our part is in order, particularly if we are tempted to see Strawson as linearly building on the work of others. I want to mention this at such an early stage in our analysis because unlike the approach of the historical figures referenced by Strawson we moderns tend to think in terms of *how* experiential phenomena could be physical in the way asserted at the start of Strawson’s argument. This is a natural result of our educational systems. In the 17th-18th centuries, however, people were not, of course, necessarily thinking in such terms as nerve connections, biological molecules, subatomic particles, et cetera; at the time to think anatomically about such matters would have involved the assumption of a direct cause-effect relationship, perhaps along the lines of how Descartes pictured the mental interacting with the physical, or perhaps employing another model. For us today then, and this is worth remembering, although when we do think on these matters we still do so in terms of direct cause-effect relations, our more advanced anatomical knowledge leads us to automatically (and therefore pre-reflectively) reduce everything to the minimal levels just mentioned. Nevertheless, even at these core and root subatomic levels, we still find that we seem to have learned almost nothing about the workings of experiential phenomena in the ‘mind’ sense that Strawson equates them with (and then additionally equates them and all else with the physical; his is a monist account).

Here already we have a first hint at just how troubling the problem of consciousness might be – and Strawson does take experience to equal consciousness to equal mind –, perhaps

³²⁶ Galen Strawson, ‘Realistic Monism: Why Physicalism Entails Panpsychism’, in *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?*, ed. by Anthony Freeman (Exeter, DEV: Imprint Academic, 2006), pp. 3-31.

even more problematic for us now than for our forebears since we find that we are bound to regard the physical in terms of the aforementioned physics-based entities and rules. This default empirical presumptiveness on our part, which practically speaking amounts to a reluctance to make (or to even really consider) metaphysical moves, causes the job to become more difficult through its shrinking of the available toolkit. Towards this problem we might therefore take a hint from the previous chapter and try instead to think of experience in a Brentano-Husserlean model, one where each experience *of* something has its intentional *for* that is immanently mental and mentally directed – even if it is also about an object ‘out there’. We might, and thinking that way might help, but that response is not Strawson’s. He does anticipate this difficulty, but rather than follow Husserl he chooses to deal with it thusly:

I am happy to say, along with many other physicalists, that experience is ‘really just neurons firing’, at least in the case of biological organisms like ourselves. But...I certainly don’t mean that all characteristics of what is going on, in the case of experience, can be described by physics and neurophysiology or any non-revolutionary extensions of them...experiential phenomena ‘just are’ physical so that there is a lot more to neurons than physics and neurophysiology record (or can record).³²⁷

One aspect that can be understood about Strawson’s account from this passage is his evident assertion that there is a ‘something more’, whatever that may be, to the physical than our empirical sciences currently teach us, and that due to that ‘something more’ experience *is* physical, whatever that in turn might imply beyond ‘just neurons firing’. In his taking of this stance we have apparently arrived at the point of departure from the standard model, the exit from which Strawson sets out in the making of his panpsychist

³²⁷ *ibid.*, p. 7.

claims, drawing conclusions that very much reject the contemporary mainstream explanatory push, as we shall see. To highlight, the pertinent points thus far are: For Strawson experience = consciousness = mind = physical (this is more or less uncontroversial); and further physical = what our standard physics says + something else (to be determined; and this is controversial).

In our earlier considerations of consciousness it was related that the present consensus view amongst cognitive scientists is one that takes consciousness as an emergent phenomena, one wherein mind arises from brain.³²⁸ Strawson rejects this quite heartily. To him, for experience to arise out of a deeper set of non-experiential (non-experiencing) elements (i.e. the fundamental elements of reality, whatever they may be; Strawson here calls them ‘ultimates’) is impossible because for something to emerge from something else (Y from X) the emerging element (Y) must be wholly dependent on the originating element (X), like the way that liquidity emerges out of its constituting molecules. In that case nothing ‘extra’ comes into the picture: liquidity is simply a result of the molecules being put together in a certain way. All of the features of Y, on Strawson’s understanding of emergence, must be intelligibly traceable back to X (and Strawson adds that he means ‘intelligible’ in a metaphysical rather than an epistemic sense).³²⁹ On this reduced notion of emergence, where nothing additional can be brought in (one could frame it as the whole must be the sum of its parts, or anyway that the parts must be ‘intelligibly traceable’ from the whole), Strawson explains both what he thinks panpsychism entails and how it might work using a more limited sense of emergence that is akin to the liquidity analogy. According to Strawson, then, in order for experience to exist – and we know with absolute certainty that it exists because we have it – experience must be an intrinsic part of every

³²⁸ It should be noted that on the consensus view, as with Strawson, consciousness is equated with mind.

³²⁹ Strawson 2006, *ibid.*

‘ultimate’ in the universe (every ‘fundamental’ or ‘absolute core particle’ or ‘base building block’, or however one wishes to label that which one cannot go beyond); and moreover that for experience to exist there must be an experiencer, a subject of experience, and hence there are truly very many subjects in the universe (as each fundamental/absolute core particle/base building block, et cetera, would qualify as an experiencer, and therefore too as a subject – here we see quite vividly the connection between his panpsychist arguments and his self concept arguments).

Strawson does not, however take this to mean that objects like stones and tables are subjects of experience (he states that such thoughts receive no support from this argument and he ‘do[es]n’t believe this for a moment’), meaning that he faces the task of both explaining – without making use of emergence as it is standardly understood – how experiencing fundamental elements could lead to both non-experiencing and experiencing matter, and also responding to William James’ well known objection of how it might be that many small subjects of experience (‘ultimates’) could possibly add up to a single large subject (you or I).³³⁰ To do this Strawson does actually make use of emergence, but in a limited manner and in a way that agrees with his totally-dependent way of thinking about emergence. He states that ‘ultimates’ have both experiential and non-experiential features, the latter necessary because some ‘ultimates’ become organized into complex non-experiential forms (stones and tables), and the former needed as others become experiential forms (you, me, a sea snail, et cetera). The experience of those later (combined) complex experiential forms, however, is very different from the experience of the constitutional ‘ultimates’, and so is in some sense emergent – but again, only in a ‘the whole is the sum of its parts (and nothing more)’ type of way.

³³⁰ *ibid.*, p. 26; James, *op. cit.*

Interestingly, Strawson follows this explanation with the following: ‘And just as there was spectacular enlargement and fine-tuning of non-experiential forms (the bodies of living things), so too there was spectacular enlargement and fine-tuning of experiential forms.’³³¹ What intrigues me most about this sentence is Strawson’s inclusion of ‘the bodies of living things’ into the non-experiential category, the category of his earlier example of stones and tables. To me this inclusion – clearly purposive and moreover meant to be illustrative – hints at a presumed (or assumed) dualism. I think this because it appears to pre-determinately segregate mind from body in its explicit placing of the latter in the ‘non-experiential forms’ camp, and in that I see here too an implication of an ‘owner’ type of relationship between mind and body à la Narveson, whereby the mental is considered wholly other than the physical (the so-called ‘boss’, as we noted earlier) and the mind can be fully separated from the body that houses the brain, which (by whatever means) is/gives rise to/contributes to/what-have-you the mind.

Near the end of his piece Strawson suggests in a footnote that electromagnetism may be ‘the heart of experience’ but it itself, in all its forms, is ‘just one expression of some single force whose being is intrinsically experiential’.³³² While this is suggestive of a buried and yet to be determined fully universal force (we might think of dark matter’s discovery here), something that is admittedly intriguing and a little mystically stimulating, we should not get too carried away just yet for it is only a speculative proposal and need not be taken as anything more than that. I include it here though because the idea does shed light on Strawson’s general thinking and furthermore provides a link to his later writing on panpsychism, work which we will consider next. Prior to that consideration though let us return to my suggestion of buried dualism in Strawson.

³³¹ Strawson 2006, *ibid.*, p. 27.

³³² *ibid.*, note 50, p. 27.

To claim that every 'ultimate' has both experiential and non-experiential features, and that the combination of 'ultimates' into this or that is what determines the emergence or absence of experience, strikes me as being a little too theoretically convenient for matching what we see in the world with the model Strawson wants to build rather than vice versa. After all, if every 'ultimate' is both intrinsically experiencing and a subject of experience, then even if the form of experience is very different from how you, I, and sea snails experience things (or rather, experience experience) then why not allow that stones and tables too have experience, only of another, different sort? How is it that the experience-having and subject-being 'ultimates' retain their experience-having and subjecthood when they combine into stones and tables but the resultant stones and tables have no experience whatsoever? There seems to be no good reason to deny this point, particularly given that Strawson does not appear to concern himself with a more commonsense approach to consciousness. (It could in fact fairly be said that really his denial of consciousness (= mind for Strawson) to stones and tables is the only standard or mainstream or commonsensical aspect to his argument.) I want to stress though that this is not to insist that Strawson must be wrong because naturalistic (or received) reasoning disagrees with him, far from it. Rather, I mean only to point out that by his own argumentative framing such a result as experience-free stones and tables does not seem to work, and that – given his panpsychist claims – there is more than a hint of dualism lurking in the background of the denial of experience to stones and tables, and on top of that his additional lumping of the bodies of living things into the same experience-free category.

Ten years later Strawson again defends his particular panpsychist position in an article that largely repeats the main points of the immediately above but that starts from a more

general series of declarations.³³³ As was intimated in the footnote on electromagnetism in his previous work, Strawson begins here from the premise that all matter is energy (safe theoretical ground ever since Einstein); this is the first of his identity claims which he acknowledges will not be argued for.³³⁴ Next, Strawson states that being is becoming and being is quality, and that, together, such identity claims mean that ‘experience or consciousness or experientiality’ is a kind of ‘stuff’, and that while experience must have an experiencer (the subject of experience), there need be no distinction between object and property, that what we rather have is a process phenomenon.³³⁵ Here too we can see the connection with Strawson’s thin self/minimal subject account, yet I think it important to highlight that ‘every experience has a subject’ does not in itself amount to ‘every substance has experience’, and that in turn does not substitute for ‘every substance is a subject’. That noted, we must also admit that given Strawson’s earlier paper on panpsychism and his account of the self, those three blocks of subject/substance might in fact be equivalent to him (they seem to be, though possibly ‘every substance has experience’ would be denied by him if ‘substance’ is understood in a broader sense than one of his ‘ultimates’, but if so Strawson would face the same lurking dualism criticism that was raised above since there would still necessarily be an experiencer(s) present by his own account), but regardless he appears to take that equivalency for granted and that the same conclusion applies in our thought too if we follow him. Be that as it may, I raise the issue at this point merely as a cautionary signal, for Strawson covers much ground in a brief time in this paper and we will need to take care lest we fall victim to the unestablished.

³³³ Galen Strawson, ‘Mind and Being – the Primacy of Panpsychism’, in *Panpsychism: Contemporary Perspectives*, ed. by Godehard Brüntrup and Ludwig Jaskolla (Oxford: Oxford University Press, 2016), pp. 75-112. I received a final version of this paper directly from the author prior to publication and accordingly the page numbers referred to in the following footnotes do not match with the chapter pages in the book (my version is simply pp. 1-35). I am very grateful to Professor Strawson for providing me with his work.

³³⁴ Strawson apparently sees no need to do so, or perhaps he feels that he has sufficiently covered the same material elsewhere.

³³⁵ Strawson also argues in a little more detail in this piece than his last that there is *never* any distinction between an object and its properties, that all objects simply are their properties; see Strawson 2016, *ibid.*

Strawson's next, and final, identity claim is that being is mind/experience, and that all being is for itself. He links for-itself with having a 'what it's like' (i.e. qualia, discussed below), which is also (for him) the having of consciousness (in whatever manner, even only sensation), which is the having of experience. All of these points lead Strawson to his latest definition of panpsychism: 'experientiality is all there is to the intrinsic nature of concrete reality', though he adds that, 'I'm understanding *experience* to entail *mind* (and conversely).'³³⁶ On his definition of it mind need not, Strawson tells us, involve any kind of intelligence. He thinks that Bertrand Russell and James may have had an alternative view – that mind does mean intelligence but that experience could exist without mind – behind their proposals that the fundamental nature of reality is 'sensation' or 'pure experience' (respectively), but that anyway on his view their positions are actually panpsychist and the disagreement about intelligence is only a terminological one. For my part I think the difference – and indeed whether or not there is a difference to be found – is rather more important than Strawson lets on, especially given that on Strawson's account all experience-having entities or elements are also subjects, even if those subjects would not have mind on the narrower view of Russell and James. Moreover, given James' combination objection (many small subjects into one large subject), I think that he in particular would find Strawson's view unacceptable. Nevertheless, let us continue to explore Strawson's latest version of panpsychist physicalism.

Emergence is again rejected and for the same reasons as in his earlier paper, namely that Y must be wholly dependent on X and if Y is not, Strawson argues, then Y cannot be said to be emergent from X. On further reflection on this point – here and with his previous work – I think we can conclude that this move by Strawson involves really a curious way to

³³⁶ *ibid.*, p. 7; emphases in the original.

figure emergence, because to employ the concept typically means to grant that something additional has been created or acknowledged or discovered in the coming together under consideration; for example traffic is often cited as an emergent property as the mass of cars behaves in a manner quite unlike a single car or even all of the cars in the same grouping taken individually, and that despite what it is each car in the traffic is actually trying to do (move!). Strawson does helpfully tell us again how he understands emergence, which is as constitution and not causation, once more using the analogy of liquidity and molecular combination to illustrate his perspective. Although we must admit that Strawson's objection to emergence is possibly only due to his more limited view of what the word 'emergence' means (he writes that he takes his stance to be the standard one based on his reading of the wider literature, yet it must be pointed out that on this position behavior as emergent is something that would be entirely ruled out, as would any additionally forthcoming properties, whether directly related to behavior or not), his denial of emergence is very firm and it is solely due to that dismissal, I think, that he forms the panpsychist argument that he does.

A few pages later Strawson writes that if one wishes to have a monist view (specifically, 'stuff monism' following his opening 'experience = stuff') while being a realist about experience (that is, admitting that experience exists), and also acknowledge the existence of non-experiential matter (energy) in the universe – a combination that many find attractive – then one will have to think that the single fundamental 'stuff' is both experiential and non-experiential, which is an impossibility. (It is anyway, we may add, at the very least logically troublesome and not likely to be very satisfying if consistency is a desired element.) Yet we must not forget that this very claim, that the single fundamental 'stuff' of the universe is both experiential and non-experiential, is how Strawson answered James' combination objection in his earlier paper: some 'ultimates' combined in one way

so that their experiential features emerged while other ‘ultimates’ combined in another way so that their non-experiential features emerged. For the experiential and non-experiential to be able to emerge in the way that Strawson allows for (we remind the reader that Strawson detailed his own weak form of emergence wherein all of the emergent Y’s features are ‘intelligibly traceable’ back to the originating X, as explained above) it must be the case that both are present. While it is possible that the seeming shift apparently being expressed in this work is simply a result of Strawson having changed his mind in the ten years between these two papers, in the discussion we are now considering there is no mention by Strawson of his earlier argument on this point, and all the talk of ‘features’ that marked that analysis has disappeared and been replaced by that of ‘properties’. (We will see below as well that Strawson does indeed try to keep a type of his specialized emergence.)

I am unsure how Strawson might delineate between these two terms (features and properties) in order to bring his papers into some form of alignment – or even if he would consider such necessary –, but whatever the case might be in support of the now alleged impossibility of a single ‘stuff’ being both experiential and non-experiential, he states in this work that his identity claim of ‘being is quality’ shows that an object simply *is* its properties, and that there is and can be no ‘haver’ or ‘bearer’ of said properties. Does this answer our doubts? It does not; once more it seems rather to go against the very argument he made and was described above, however for the purposes of discussion we will not dwell on this and merely take note of the contradiction without giving it weight (again, Strawson may simply have changed his view and we can be charitable on that, but then further down in this latter work he again raises (or resurrects) his kind of emergence). We will now instead proceed to consider his (new) points which criticize a monism that allows both experience and non-experience to be part of the same ‘stuff’; yet while doing so it may nevertheless be instructive for us to at least keep the position that Strawson is coming

out of (his earlier arguments) in mind.

For it to (now) be an impossibility for something to be both experiential and non-experiential rests entirely on Strawson's 'being is quality' claim (which again is not argued for in this paper), since if a 'haver' or 'bearer' is allowed then of course that 'haver' or 'bearer' might contain any number of qualities/properties/features, et cetera. Perhaps some are experiential and some not so, after all why not? Such was the stance taken previously. Moreover, it is only, I think, by defining the non-experiential in the directly negative as-opposed-to-experience manner that Strawson does that enables this line of his reasoning to work for him. Understanding that grounding to the interplay being considered, we are tempted to push against it by asking: Why not allow the non-experiential to be neutral? Why must the non-experiential be the polar of the experiential, the negative to the positive? After all, if we want to think in particle terms then experience could be a positron while non-experience is a neutron rather than an electron, or the like.

In response Strawson could well fault this kind of thinking as an inability to get around the physics-based models we use to picture the world, and while I agree that my example just now is probably too simplistic, if we are talking about the intrinsic nature of energy (= all matter) then even so I do not see why that nature must be mono-anything, nor why it must be mono-experiential simply because experience exists (allowing that it does exist, some do deny that). Moreover, although granting that we were to go along with Strawson's 'stuff monism', we would still not need to equate 'monism' with 'monochrome' or 'mono-flavored'; the same 'stuff' might be compositionally uniform but expressed differently, as with Strawson's earlier argument about 'ultimates'. To be clear, I am not agreeing with Strawson's 'ultimates' viewpoint nor with its derivative conclusions, I am rather just pointing out that though that argument did admittedly seem to be dualist it still

appeared to work much better than this later argument does, which after all relies rather heavily on declaratives.

To put all this another way, Strawson seems to presume that experience exists at a root or core level, indeed at the most fundamental root or core level (we might think again of his ‘ultimates’), and then from there that the non-experiential or the not-experiencing or the unexperiencing must either exist (if it exists) at the same level or at a level above it. He does not leave any space open for experience to be both real and non-fundamental because he takes experience to be a real quality and all real qualities to conversely be real beings (Strawson states at the outset of this piece that all of his identity claims are fully reversible). Strawson’s definitions are what really seems to be doing all the work in this argument: experience is real, therefore experience is ‘stuff’; ‘stuff’ is being, being is quality, therefore experience is quality; being is quality, therefore opposing qualities cannot be of the same being, therefore non-experience cannot be of the same being as experience; experience is real, to be real is to be, experience is but cannot also be non-experience in its being (logical impossibility); therefore no non-experience, therefore only experience, therefore panpsychism. The logic is sound but only if we accept all of the definitions on their faces (to me this is a wonderful example of much that is wrong in analytical philosophy), and there seems to be both many points that are assumed in the definitions themselves and many points about which we might hesitate to be as confident as Strawson appears to be.

The implications that this panpsychist outcome holds for consciousness, and therefore too for the self (particularly on a theory of the self like ours), are far-reaching whichever grounding is taken to be the stronger one (by which I mean that from Strawson’s earlier article or from his later one). This is especially important, I think, in regards to subjecthood and the line at which that is drawn, and for Strawson that line remains a minimalist one, as

we shall soon see.

On the note of confidence and admitted ignorance, in this piece Strawson repeatedly makes the point that we think we know much more about the physical than we actually might, that in particular we could be quite wrong about space (and therefore spacetime and therefore the very fabric of the entire universe), and so we only think we know that in accepting that conscious experiences are brain states ‘there are things that have both E [experiential] and non-E properties.’³³⁷ Again, to Strawson we think we know this but do not actually know it. Strawson replies to this reported overconfidence on our part by pointing out the presumption involved in the position that ‘a thing can’t possibly occupy space without having some non-experiential stuff being.’³³⁸ In this Strawson wishes us to free ourselves of the idea that a ‘mere object’ must be non-experiencing, and perhaps he is right that that particular idea is in error, but that is not much support for his conclusions because again we can in turn raise the same objection we did previously that if experience is an intrinsic and fundamental – solely fundamental – part of absolutely everything then why must we deny it to the kinds of inanimate objects we unreflectively refer to as ‘things’? Maybe you and I do not, but recall that Strawson did just that in his earlier work, and he does so again in this later paper too, stating that his view ‘has nothing to do with Berkeleyan idealism or phenomenalism, and it certainly isn’t committed to the implausible view that tables and chairs are subjects of experience’.³³⁹ If we are to go this described panpsychist route I must admit that I cannot see why tables and chairs (et cetera) need be excluded, regardless of the presence or absence of human perceivable manifestations of said experience in tables and chairs. Again, *if* we are to go this route, which of course we need not. On what are subjects of experience for Strawson, he here in this work re-states his ‘experience must

³³⁷ *ibid.*, p. 16.

³³⁸ *ibid.*, pp. 16-17.

³³⁹ *ibid.*, p. 20.

have an experiencer' position although he now does, in keeping with his arguments about the self which appeared three years after his earlier panpsychism paper, make the distinction that an experiencer need be neither separate from nor last longer than the experience (aligning with his thin self/minimal subject view).³⁴⁰

In analyzing this aspect we need to remember that by Strawson's own definitions experience = consciousness and therefore for him the having of experience = the having of consciousness; in stating so flatly that the inanimate do not have experience Strawson must then at least be holding on to emergence in the experientially manifesting way that he did in his prior paper given his panpsychism. He seems to want to have experience be everywhere and a part of everything but not so much so that it shows up in 'mere things', because that would require, on his view, that those same things also have consciousness, and how can stones, tables, paper clips have consciousness? (Others do make that case but Strawson clearly rejects it.) In this later paper as well he in fact tries to admit some forms of emergence and deny other forms in a way that fits this split experiential world (i.e. Strawson's stance that all matter undertakes a journey from a basic and intrinsic experience to sometimes exhibiting experience and sometimes not) by repudiating what he calls 'radical emergence'; in the earlier work such was labeled 'brute emergence' but definitionally they are the same (the liquid analogy mentioned above is used in both papers). This however raises the same questions for us that we already considered, questions about the denial of a more standard take on emergence and about the limited way in which he defines emergence such that he can neatly differentiate between his acceptable emergence (of the constitutional type) and what he calls 'radical' or 'brute' (of the causal type). A larger point here though, is this: in a response to Strawson's earlier work Daniel Stoljar pointed out that Strawson appeals to ignorance to make his arguments,

³⁴⁰ *ibid.*, p. 19.

such as in the above where Strawson critiques our presumption in knowing about space, yet also sometimes rejects appeals to ignorance, such as with emergence and the idea that emergence might be what is happening and we simply do not understand how it is happening.³⁴¹ Strawson addresses this issue in a footnote by summarizing it and then dismissing it simply with, ‘This is true but I do not think a difficulty.’³⁴² For reasons of a want for consistency and to avoid what appears to be cherry-picking on Strawson’s part I find Stoljar’s criticism to be valid, and given that emergence in many ways works much better than panpsychism, or at least matches much better with our empirical sciences (and commonsense or intuitive position), I think it deserves more of a response than Strawson gives it – at the very least in explaining why he does not think it a problem since he himself appeals to ignorance in defending his more limited type of emergence, not to mention his felt need for some form of emergence to begin with.

Strawson writes:

Call these microexperiential phenomena [of subatomic, atomic, molecular, cellular energy] *Es*. *Es* may have phenomenological features of which we have knowledge, in having the kinds of experiences we have, even though they somehow conspire to constitute our experiences. The energy that is an electron is wholly a matter of experiencing, on the present view, but the specific phenomenological character of this experiencing may be radically unimaginable by us.³⁴³

We may concur with Stoljar that Strawson seems to be choosing fairly loosely and freely

³⁴¹ Daniel Stoljar, ‘Comments on Galen Strawson’, in *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?*, ed. by Anthony Freeman (Exeter, DEV: Imprint Academic, 2006), pp. 170-176.

³⁴² Strawson 2016, *op. cit.*, p. 17 note 30.

³⁴³ *ibid.*, p. 25.

what is unproblematically unknown and what is problematically unknown, and again how such a view should lead to panpsychism rather than to ‘radical’ (i.e. a more standard) emergence is not clear. In both cases we must admit that we do not know the exact *how* of it but we can see the results; which path appears more simple, elegant, and attractive? On the one hand we have a certain unimaginable form of experience as the fundamental and sole intrinsic nature of all energy (which is all matter), but that only sometimes becomes through that same energy another form of imaginable (to us) experience depending on how the energy/physical elements involved in the resulting chunk of matter have been combined, while on the other hand we have fundamental energy in the way that physics tells us (investigations into which are ongoing) which is non-experiential but that also sometimes combines into chunks of matter (like us) that have experiences which we can recognize. The choice when put in this way is between the panpsychic positing of experience (understood as consciousness) all the way down, or the emergent positing of experience (not necessarily equated with consciousness) at the observable level: either way there are parts we do not know.

Although I think that from the agnostic or ignorant position we find ourselves in Strawson’s complete rejection of so-called ‘radical’ or ‘brute’ emergence seems overly hasty (whether or not one wishes to tack on to that his embrace of panpsychism), particularly given just how much he is willing to admit that we do not know, other problems start to mount for his argument as it goes on, and these too connect with the self and/or subjecthood in the descriptive reaches concerned. The reader will recall that one of Strawson’s primary identity claims is that ‘being is quality’, and that it is due to that identity claim that he builds the panpsychist case that he does. Nevertheless, that very identity position appears to be weakened by a (yet another) later appeal to ignorance that Strawson makes when he writes that, ‘The point is not just that the numbers and equations

of physics don't capture the *whole* basic or essential nature of reality; it's that they tell us nothing about the nature of concrete reality insofar as its nature is more than its structure.'³⁴⁴

The 'numbers and equations of physics' are mathematical representations or descriptions of observed/observable qualities, properties, and functions in the natural world; that much is surely agreeable upon. Yet if being is quality, and therefore – as Strawson repeats often – there is no distinction between an object and its properties, then something here is amiss as Strawson's complaint with physics would seem to imply that being might be more than its quality(ies) if an object/entity's nature is 'more than' its structure. How can being (intrinsic nature) be quality undistilled from its holder without the descriptions of that quality telling us about the holder? Moreover, if being really is quality then how could an object's nature be more than its structure anyway? (This last question seems particularly apposite given Strawson's rejection of all but his own type of emergence.) In a footnote on the same page Strawson does say that unlike Russell he takes a 'concrete thing's structural nature to be part of its intrinsic nature',³⁴⁵ yet that does not seem to really clear up the picture for us as if that were indeed the case, and physics were telling us something about said structure, and further if the claim 'being is quality' is true, then surely we would at least have *something* about (instead of 'nothing about') the nature of concrete reality even if said structure were just 'a part of' (and not the whole of) said nature. At any rate, in their repeated and varied usages by Strawson there appears to at least be some terminological overlap happening between the terms 'thing', 'object', 'stuff', 'being', 'concrete reality', 'essential nature', 'intrinsic nature', 'quality', 'property', and probably some other labels that I have missed (e.g. feature?). Such might account for the perceived problems here, but

³⁴⁴ *ibid.*, p. 26; emphasis in the original.

³⁴⁵ *ibid.*, p. 26.

it is difficult to see how more transparency could be achieved with this conceptual framework just as it is given.

As for James' combination problem, which Strawson considered a central and serious issue in his previous paper, in this latter piece he states that he has lost all sense that we have any reason to think it an obstacle and that he 'can't feel' any deep difficulty in the objection. Strawson concludes this apparently based on what he determines we have reason to expect, and that we cannot really suppose whatever 'Laws of Experiential Combination' might be operating to be any more clear to us than those of quantum mechanics. He writes that when he puts his mind to all that he does not know and understand about quantum entanglement, synchronic/diachronic identity, field theory concepts, et cetera, he does not see how not knowing about how trillions of experience-having subjects can combine into a single experience-having subject is an issue, but he does think that 'merging droplets of water provide a nice image: seemingly sharply bounded individual items that merge on contact into one equally well bounded item'.³⁴⁶

In effect I think that Strawson's argumentative move here is essentially to say that given everything he does not understand he cannot see why the combination problem is a problem, and so he dismisses it as a problem. That may be fair enough for what it is, but his analogy with merging water droplets is still somewhat misleading as what they represent is a structural merger or combination, not untold independently experiencing subjects becoming a single – and singularly experiencing – subject. Even if we were to grant that each water droplet has its own 'mind' of a sort (somehow, though it too would be a compositionally mega-subject if all its constituent fundamental particles or 'ultimates' or energy bits are each minded subjects as per Strawson), those droplets coming together into

³⁴⁶ *ibid.*, p. 28.

a greater body of water would not have the same ‘mind’ that we do, and that judgment would apply even given a full-scale acceptance of Strawson’s account, although he denies mind (= consciousness = experience for him) to the inanimate, as above. (Each ‘minded’ ultimate would go towards forming a non-‘minded’ water droplet which would go towards forming a non-‘minded’ greater body of water; the analogy is not about multiple experiencing subjects becoming a compositionally unified and singly experiencing subject.) Furthermore, although he does not state this outright given the narrow panpsychist focus of the article, we would have to think that in denying mind in this way Strawson is also necessarily denying subjecthood because for him such is based on the stance that all experience has an experiencer (again, the thin self/minimal subject). The picture of how the universe might very well work that emerges (for we accept emergence!) from these reflections of ours is far more in line with the standard views that take the having of experience to be grounded in brain-based living organisms than it is with Strawson’s panpsychism.

Strawson closes his essay with another charge against emergence, arguing that all non-panpsychist hypotheses fail because they require ‘radical emergence’ at some point and that the experiential simply cannot emerge from the non-experiential. He writes that, ‘Long familiarity with a picture according to which experientiality emerged from non-experientiality in the course of biological evolution has softened our thinking in such a way that we can no longer clearly see what an extravagant hypothesis this is’.³⁴⁷ It is helpful and instructive that Strawson brought up evolution as he did because evolution is a good example of how we are sometimes quite wrong (and/or unimaginative) in how we picture the world around us. (On this we can agree with Strawson that there is so very much we do not know.)

³⁴⁷ *ibid.*, p. 32.

Although biological evolution is indeed thought to be a usually incremental process it can sometimes happen in great, inexplicable leaps such as was reported about the ‘lightning fast genetic evolution’ of Tasmanian Devils after a facial cancer epidemic threatened the species with extinction. In just four to six generations scientists discovered a *species-wide* adaptive change to seven genes in the Devils’ genome;³⁴⁸ the specialists involved were shocked at the breadth and rapidity of the evolutionary response. While it is true that evolution is of course not emergence, if a process like evolution can see its own sudden jumps (which Strawson denies occur in nature as part of his attack against ‘radical’ emergence) then it might be that emergence of the type that could lead to experience out of the non-experiential and/or consciousness out of the non-conscious is also a possibility. Perhaps; I think the door can safely remain open. How it is a possibility we admittedly cannot say, but prior to Mendel, Watson, and Crick we would not have been able to ‘intelligibly trace back’ (as per Strawson’s demand) the Devils’ evolutionary response although we had Darwin’s theory firmly in hand. That at least indicates, I think, that Strawson ought to pause before he so firmly rejects emergence, even if he does for other reasons prefer panpsychism. After all, as he has shown himself willing to take an agnostic stance in various other areas then why not on this point too? For the rest of us who may not be as (or at all) enamored with panpsychism, it also gives us reason to continue to look at emergence as one option for how the puzzle of consciousness – and the question of experience – might be solved. Here too I believe it best to remain willing to entertain many different kinds of emergence, be they ‘radical’, ‘brute’, ‘limited’, et cetera. Any progress on this point would also inform the conclusions we might draw about the self.

³⁴⁸ AFP-JIJI, ‘Tasmanian devils pull back from brink of extinction’, *The Japan Times*, 24 September 2016. <<http://www.japantimes.co.jp/life/2016/09/24/environment/tasmanian-devils-pull-back-brbri-extinction/#.WEpn8LlqMuc>>.

B. Responses to Strawson and related panpsychist concerns

We will now turn to some of the responses that have been made to Strawson's argument for panpsychism in order to fill out our understanding of it and of the related points dealing with experience and consciousness, all of which should shed more light on the self. The replies to Strawson in the below were made to his earlier of the two papers outlined above, but I believe they are still applicable to both of his pieces given that Strawson's later article largely repeated and only slightly expanded on his earlier one. We should note that many of the responses in what follows grant or perhaps accept his equating of experience and consciousness; while this is more of an issue for the next chapter, I would still like to point out again at this juncture that the equation is not without its faults and limitations. By stating that experience = consciousness Strawson is blocking the unaware (i.e. the preconscious mind) from being involved with both sides of his equals sign. It has moreover been argued in this study that perhaps we ought to consider ourselves to always be conscious in at least some way if we accept the current view prevalent amongst cognitive scientists, that is, the view that the brain's networked structure and cross-communicative nature is what gives rise to the sense of mind or unified consciousness with which we are familiar. If at a minimum one of those networked sections is working then, or so we have put forward, one is conscious. If we do take this stance it means further that even while asleep we are conscious, and that if that is the case then to 'be conscious' in the technical sense does not have the 'be cognizant of' meaning that the same phrase does in the type of everyday language that Strawson and many other thinkers about mind tend to use when they discuss consciousness. As noted, this is very important for our view of self since consciousness is one of its primary constituting elements. Let us therefore determine what more might be learned about consciousness from these responses to Strawson's ideas, and thereby too what might be learned about the self. First though a few words on this parsing of 'consciousness'.

The nuance between ‘be conscious of’ in our technical sense versus ‘be cognizant of’ in the less analyzed sense is an important one, and it in turn means that experience might be separated from consciousness by that very red line (if we wish to call it that) of awareness. To have an experience is to be aware of that experience and to be aware of the having of that experience – a point Strawson himself makes –, but I would argue that that awareness itself does not necessarily equal the entirety of consciousness.³⁴⁹ As suggested in previous sections of our study, we might speak more clearly about such matters if we employ ‘consciously aware’ or even just ‘aware’ instead of the unreflective usage of ‘conscious of’ (while still maintaining the same sense), and then use ‘conscious’ or ‘consciousness’ to only be about the networked system as such, and not therewith to necessarily imply any degree of awareness. Whether we agree with the current cognitive picture or not (the idea that we are always conscious in the way described does appear to be gaining more ground³⁵⁰), I think that the unestablished and non-argued for equating of the two terms is likely to lead to confusion, and this is the case even if the converse stance to the above is taken and experience is instead allowed to be aligned with the unaware and to be connected to such mechanisms as the preconscious perceptions that generate the many intuitive decisions and judgments that mark our days but which do not enter into our conscious (aware) recognition. This matters because what we choose to take as our background position(s) will determine which side we think can transcend awareness, either consciousness or experience, or indeed that neither side can transcend awareness and that consciousness as a term only applies to instances of awareness (and presumably therefore the terms ‘unconscious’ or ‘subconscious’ would be marked off by a thick division from

³⁴⁹ Here we may wish to remember too Strawson’s thin self/minimal subject argument and that for him subjecthood, consciousness, experience, are all mere instants – subfractions – of a moment.

³⁵⁰ See, for example, the work of Giulio Tononi. A summary of his ‘integrated information theory’ can be found here: Giulio Tononi, ‘Integrated information theory’, *Scholarpedia: The peer-reviewed open-access encyclopedia* 10:1 (2015), 4164. <doi: 10.4249/scholarpedia.4164>.

‘conscious’).

We now embark on our consideration of replies to Strawson’s panpsychist position, beginning with Sam Coleman. Coleman points out that Strawson’s panpsychism suffers from its own unexamined physical assumption in its ‘smallism’, the view that there are layers to reality and that the upper layers are built from the smaller ones below. Coleman states that there is an alternative physicalist position, ‘macro non-reductive physicalism’, which (may) hold that layers of reality can be ‘closed’ such that the mental, say, just is a closed layer that appears at certain levels of organization.³⁵¹ Coleman does not expound much on this idea of layers, but it appears to be based on emergence, with a particular complexity first being required; without knowing exactly how such a process might work we can perhaps leave it at being an interesting idea. ‘Smallism’, however, does seem a clear product of the physics-bound thinking that we in modern times are prone to (and which Strawson is at pains to criticize), yet without some substantive reason for giving it up I do not see why such abandonment might be called for whatever the theoretical complications of holding onto ‘smallism’ are, especially given the overwhelming empirical basis of an idea like ‘smallism’. There is a sense in which the (somewhat) kneejerk reaction to apply reductivism can certainly be misleading, but there is also the sense in which the concept has gained such depth and stability because it seems to work so well. Explanatory potency born out in multiple experiments has made ‘smallism’ a staple of science – and really, given all that, why should it not be? Panpsychism may be trying to blend physics with metaphysics, but it still purports to speak for the empirical world (and even to do so ontologically), and so certainly such might apply. Yet also to the self? That, I think, depends entirely on the self theory in question: for ours, with its basis as a naturally

³⁵¹ Sam Coleman, ‘Being Realistic: Why Physicalism May Entail Panexperientialism’, in *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?*, ed. by Anthony Freeman (Exeter, DEV: Imprint Academic, 2006), pp. 40-52.

occurring and ongoing ‘psychological unit’, relations and connections between its elements can easily be found, and we need not on the face of it feel compelled to dissociate with ‘smallism’ either. (Although if well argued I think we might and still not lose the essentials of our concept of the self, including its personal identity and whole person aspects, particularly if the ‘closed layers’ involved are in fact bound by emergence in the way they appear to be.)

Regardless, Coleman does not dwell on the issue himself but instead turns quickly to another critique of Strawson, namely that he is a ‘microproperty dualist’ in his description of his ‘ultimates’, some of which, Coleman asserts, have experiential and non-experiential properties. Coleman writes that, ‘Given the difference between experiential and non-experiential modes of being – the fundamental gulf in ontology for Strawson – the accompaniment of one property type by the other at the lowest level cannot be other than brute: there can be no reason in the experiential nature of an ultimate why it is also non-experiential, and conversely.’³⁵² Although the charge of being dualist is a fitting one, and one that we too made earlier, I think that in fairness Strawson’s position is such that *all* ‘ultimates’ have *both* properties but only at certain organizational levels does experience manifest itself according to Strawson’s particular understanding of emergence, something which Coleman either glosses over or ignores in his response. Instead of panpsychism Coleman suggests ‘panexperientialism’, whereby the same ‘ultimates’ that Strawson treats have ‘feel’ but not representation, and where subjecthood is limited to organisms possessing a brain. In such a picture, Coleman states, ‘ultimates’ are ‘intrinsically conscious’ but may or may not be so placed as to represent.³⁵³ Again, the idea is interesting, and the step back from the ‘every “ultimate” is an experience having and

³⁵² *ibid.*, p. 43.

³⁵³ *ibid.*

conscious subject' stance is a welcome one, but just what Coleman means by 'conscious' is left open and we still lack any good explanation for the why of it, not least the how. He has thankfully moved us away from Strawson's picture of subjecthood, however, and closer to our own view of a self needing a brain (in our terms C + B_p (consciousness plus bodily presence, with consciousness emerging out of a brain)). While those terms are not equal – I might be tempted to equate subjecthood with personal identity on our scheme, though it probably fits better at the whole person level – Coleman's analysis is still informative for us regarding issues related both to consciousness and to the self.

Philip Goff is more directly critical of Strawson's position, starting from his analysis of the 'knowledge gap' upon which Strawson assumes that we have fully transparent access to the true nature of conscious experience (and Goff's use of 'conscious experience' is far more in line with my thoughts above than the usual running together of both terms into a single one is), but that that itself – the knowability of the mental – is an epistemological issue. Goff suggests that we might not know, and that what we do not know (that is, what is conceptually distinct in physical and functional properties) is nevertheless empirically identical.³⁵⁴ In our incomprehension we perceive a difference where there is none. Strawson would no doubt object to this line by admitting that he concedes we do not know much, but Stoljar might applaud it, for again Strawson does seem to pick and choose when ignorance helps his cause and when it hinders it. Goff continues his examination of the purported transparency of introspection by asking how it could be that, if indeed the mental/experiential is transparent, we yet experience our trillions of individual subjects and separately experiencing 'ultimates' (as per Strawson) as a unity that appears to us as being unbreakable? This is a valid criticism, I think, and much related to James' historical

³⁵⁴ Philip Goff, 'Experiences Don't Sum', in *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?*, ed. by Anthony Freeman (Exeter, DEV: Imprint Academic, 2006), pp. 53-61.

objection and the question of self-identity. Goff also (with us) points out that Strawson's favored form of emergence (intrinsically experiencing energy or 'ultimates' into sometimes experiencing more complex forms and sometimes non-experiencing more complex forms) is no different than the emergence principles used by standard *a posteriori* materialists. I agree here too with Goff, and would like to stress once more that (standard) emergence also appears to be a far simpler and more elegant solution than the positing of experience (and certainly more so than the positing of subjecthood) onto the fundamental constituents of our universe (e.g. 'ultimates'). With Goff's critique we gain support for understanding consciousness as both life-bound and as emergent, a helpful step in our examination of it and in how it relates to the self.

Frank Jackson makes some particularly astute remarks on the assumptions buried in Strawson's thought, noting that extra properties that appear from aggregates (e.g. power from an engine) supervene on the parts involved in a kind of *a priori* emergence (the emergence being determined *a priori*);³⁵⁵ the implications of this seem especially apposite towards Strawson's later defense of panpsychism given his foundational identity claim there of 'being is quality' (hence, per Strawson, there is no distinction between an object and its properties). Would power be an inherent – compositionally, definitionally – property of each object that combines to form an engine? It would seem that for Strawson it would have to be, yet in what way is hard to imagine without adding all sorts of highly complicated quality properties to each object/being that one comes across. (We are immediately tempted to ask questions like where the power is to be found in a nut lying on the garage floor.) Jackson further remarks that there are many types of consciousness, and many differences therein that are related not only to situation or neurology but also to such

³⁵⁵ Frank Jackson, 'Galen Strawson on Panpsychism', in *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?*, ed. by Anthony Freeman (Exeter, DEV: Imprint Academic, 2006), pp. 62-64.

details as blood sugar levels. If various types of consciousness can be formed by the same items, he asks, then why not consciousness itself as well?³⁵⁶

It is a good question and raises an issue that has begun to receive some attention: that of the character and types of mental phenomenal experience. Uriah Kriegel has outlined that there are at least six distinct forms of cognitive experience: thought and judgment, will/agency, apprehension, emotion, moral thought/experience, and the experience of freedom.³⁵⁷ Here we need not delve into any of the intricacies involved for it seems sufficient to once more merely highlight the somewhat clumsy results that appear when experience is equated fully with consciousness. We may in addition ask about instances where one has a cognitive experience and is not even made aware of it (e.g. a flitting glimpse of something that failed to register in one's active attention but that still generated a brain signal and/or other automatic processing response); such occurrences nevertheless affect behavior in the influence yielded on the intuitive judgment system and its corresponding outputs, and thereby in turn inform the self via that internal data set. Would an unaware cognitive experience be an example of a conscious experience? What could be said of its phenomenology? Strawson's account is silent on such matters, yet on its face we might think that his model would hold that the 'ultimate' subjects involved each had that experience, or some form of it or some form related to it, and that the entire episode would be 'conscious' (equated with awareness) in some manner. A moment's reflection however shows that this is clearly not satisfactory at least when it comes to the everyday, wherein we function as a single unified subject. In its failure to distinguish between experience and consciousness Strawson's panpsychism seems at a loss in dealing with many of the complications of actual subjecthood that arise at the felt and lived human level, even if that

³⁵⁶ *ibid.*

³⁵⁷ Uriah Kriegel, *The Varieties of Consciousness* (Oxford: Oxford University Press, 2015).

level was not Strawson's aim in writing his piece. It is perhaps unfair to blame Strawson for avoiding such considerations as his objectives were elsewhere, but we can find fault with the model for its apparent inability to be applied to such questions, and with that we once more move forward in our consideration of consciousness' grounding.

Colin McGinn also approaches the question of panpsychism and the problems that it attempts to deal with from the angle of all that we do not know, suggesting that given our great incomprehension it may be too simple to think in terms only of the experiential and the non-experiential, and that out of our epistemological blanks regarding matter and consciousness a solution that could explain emergence might come.³⁵⁸ It does seem best to keep all options open, as mentioned above, and McGinn's querying of the dichotomy Strawson has set up has the added advantage of highlighting again the inadequacy of the oppositional way that Strawson defined his experiential properties (either fundamentally completely present or fundamentally completely absent). McGinn also offers as a possible analogy the problem of joining classical physics with quantum physics: there too the one somehow emerges from the other but we do not understand how.³⁵⁹ We might similarly look to life itself for a parallel: by Strawson's argument it would seem that we would have to admit that we know life, know it fully and incontrovertibly, that we have no (inside) evidence for non-life, that because life is it must be the intrinsic nature of all energy/'ultimates', and that aside from structural issues biology has nothing at all to say about it. Our bodies are composed of living cells, of course, but what about the proteins and molecules making up those cells? What of their atomic particles and subatomic particles? Does 'life' exist everywhere and at each level? If we accept 'smallism', as

³⁵⁸ Colin McGinn, 'Hard Questions: Comments on Galen Strawson', in *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?*, ed. by Anthony Freeman (Exeter, DEV: Imprint Academic, 2006), pp. 90-99.

³⁵⁹ *ibid.*

Strawson does, then we quickly find ourselves having to apply the ‘pan-’ prefix to all sorts of observable phenomena once we reject emergence as an explanatory option. This is a crucial point, for not only the desirability but also the necessity of emergence is easy to overlook. Nevertheless, the exaggerated ‘pan-’ affixation is thankfully, reassuringly, not needed if we accept both ‘smallism’ and emergence, and so here too consciousness as a property that arises out of a functioning brain presents itself as an attractive theoretical perspective, and with that its position as a constituting element of self.

Presenting a more solidly materialist stance, David Papineau takes it that given sufficient advances in brain science we will someday be able to establish that phenomenal kind M is identical to physical kind P, and that such M = P identities will ‘fully capture the nature or essence of experience’ (quoting how Strawson put it to make the opposite point) in physical terms, ‘that the relevant physical term will refer to nothing other than the phenomenal kind M.’³⁶⁰ As a ready example of where just such a trend has occurred (and regularly occurs), Papineau writes that, ‘Many true identities are only established a posteriori, such as that table salt is sodium chloride... But that is no reason to deny, say, that “the term *sodium chloride* fully captures the nature and essence of table salt”.’³⁶¹ This notion of physical terms I find highly instructive; could such thoughts not also be combined with ideas related to qualia (to be discussed at more length below)? Must a relationship defined, in physical terms, by M = P also define the ‘what it is like’ aspect of the same phenomenal kind? I can see no reason why it should need to – though I can see why we might wish it to –, particularly given that we cannot even define in phenomenal terms ‘what it is like’ for each other (apart from in broad and inexact stabs at it, that is). My

³⁶⁰ David Papineau, ‘Comments on Galen Strawson’, in *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?*, ed. by Anthony Freeman (Exeter, DEV: Imprint Academic, 2006), pp. 100-109 (p. 101).

³⁶¹ *ibid.*, p. 102.

sadness is not yours, and I will never know how it feels for you.³⁶² May we not therefore think a physicalism that is materialist in the (standard) way Papineau describes as being one that is compatible with a strong emphasis on qualia and phenomenal kinds in general?

Papineau furthers his focus on terms by explaining that he has no difficulties in accepting scientific phrases as able to ‘fully capture the nature or essence of experience’ because he does not recognize any way in which the mind ‘captures’ something apart from simply referring to it, and he thinks that our common intuition of a distinction between brain and mind is the culprit behind our confusion on this matter. Papineau argues that the stance that consciousness is completely transparent to phenomenal thinking (and thus not reducible to opaque physical states) – a stance Strawson not only takes but exhorts – stems from the introspection showing, ‘that conscious states are present in our phenomenal thinking, in a way that the referents of other thoughts are not’, that ‘the conscious *referent* itself is involved in the *vehicle* of thought’.³⁶³ Papineau calls such instances a ‘use-mention feature’ and thinks that it is this characteristic which is responsible for our unreflective tendency to separate brain and mind, an intuition that he thinks we ought to drop if the straightforward physical causal-explanatory arguments point towards standard (opaque) physicalism of the table salt = sodium chloride kind.

We will be considering the neuroscientific and cognitive accounts of consciousness in the next chapter, and we will certainly find similarly robust materialist claims being made there, but just from Papineau’s argument here I think it should be becoming increasingly clear that one of the major obstacles we face in trying to solve the perceived difficulties of

³⁶² We are reminded here of the full ramifications of the famous ‘problem of other minds’ and the necessary epistemic blank that is entailed in my attempts to truly know you, and this of course all the more so when it comes to interspecies speculations and not just intraspecies ones.

³⁶³ Papineau, *ibid.*, pp. 104-105; the first quoted section is from p. 105 and the second from p. 104; emphases in the original.

consciousness and/or experience is in how we have been *discussing* these problems, the actual terms employed and therefore the manners of thinking generated by those terms. Of interest too, and also for further reference, is the notion that our qualia, our sense of ‘what it is like’ may simply be a physically built-in feature of how our brains work and carry no further implications beyond that. Everything about our mental lives may indeed be reducible to neurons, particles, and the connections in between, as incredible as that strikes us looking out at the world with our view from the inside. Might the self too, even as a ‘day-to-day psychological unit’ as we have described it, also have physically traceable roots? Any conclusions on these issues seem premature, but they are intriguing.

Taking the topic in an entirely different direction, William Seager writes that physicalism may be able to account for consciousness and experience by first rejecting the ‘smallism’ or ‘reducibility principle’ that Strawson adopts and instead asserting that all there is to matter are the relationships that science shows; this outlook he calls ‘relationalism’. If we adopt a ‘relationalist’ stance, Seager argues, it becomes possible to say that what is lacking in our current explanations of consciousness/experience is discoverable by science for ‘there is nothing more to consciousness than its place in a system of relations linking it to events in the material world as well as other mental states.’³⁶⁴ Seager’s ‘relationalism’ moreover admits of hierarchies of the relational structures (although without using a ‘smallist’ basis as the relational system itself is enough to ground the reality of the entities involved (thus differentiating it from functionalism)), but it is ‘relations all the way down’: meaning that for consciousness, ‘its proprietary system of relations might well be a neural structure’.³⁶⁵ In the latter half of his essay Seager goes on to consider some arguments

³⁶⁴ William Seager, ‘The “Intrinsic Nature” Argument for Panpsychism’, in *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?*, ed. by Anthony Freeman (Exeter, DEV: Imprint Academic, 2006), pp. 129-145 (p. 139).

³⁶⁵ *ibid.*, p. 139.

against ‘relationalism’, and although due to the structure of his piece it is not clear just what he wishes to advocate – perhaps he only wishes to consider the varying sides – he appears to find that the only real counter to ‘relationalism’ would be something along the lines of an intrinsic nature position such as Strawson’s (or others’). His explication of the intrinsic nature position is revealing though, and it does warrant some scrutiny.

Seager brings up the ‘evidence that some things (such as *me*) have intrinsic properties’ and suggests the idea that given phenomenological data it looks apparent that ‘my current state of consciousness seems to be something that could exist even if I were the only thing in the universe’, and further that ‘Consciousness itself provides perhaps the best argument that there are intrinsic properties, and this is exactly why consciousness appears as so troubling or even alien to the scientific picture of the world which deals exclusively with relational or structural features, leaving aside any attempt to grapple with intrinsic natures.’³⁶⁶

I find this line of thought regarding consciousness to be fraught with conceptual confusion. To begin with, regarding the purported evidence that one’s current state of consciousness could be alone in the universe, we must ask what the ‘I’ is that is holding said consciousness. Even if that ‘I’ is alone in the universe that ‘I’ must be embodied in some manner else ‘I’ simply equals ‘just consciousness’, but then whence the subject? We would have to venture into either something like Nagel’s suggestion that consciousness might be a fundamental building block of the universe (his take on the intrinsic nature stance, consciousness as a potentially ubiquitous property)³⁶⁷ – yet if so where is the specific and subject-bound ‘my current state of consciousness’? –, or we would have to make our way to something like Strawson’s own thin self/minimal subject account where consciousness

³⁶⁶ *ibid.*, pp. 143-144.

³⁶⁷ Nagel 2012, *op. cit.*

is experience is subjecthood ('being is quality') without needing any bearer whatsoever – again begging the same 'my current state of consciousness' subject question (and resurrecting James' combination objection). At heart here is the rational factor that we can only begin to think about consciousness in such subject-absent or subject-disregarding ways if we first assume some kind of (possibly latent or unrecognized) dualism, either of a property or a substance type. If consciousness cannot be separated out from matter, which is after all what physicalism mainly claims, then we cannot consider it outside of relational or structural areas; to be disparate it would have to immaterially float 'out there' which cuts against all of physicalist theory. We would have thought without a thinker. Ignorance of the physical universe could be appealed to here, it could be argued that we simply do not really know what the physical is (Strawson frequently asserts this), but that is little better than a place holder argument with less grounding than the opposing (and better logically supported, I think) epistemological blank argument. Those who take the latter and put their hope in science someday filling in the open spaces in our understanding are affixing their trust to an ongoing investigative system that is organized and procedurally executed. Those who take the former and put their hope in the thought that we know nothing and therefore ought to consider everything regardless of prior basing or even of seeming sensibility are affixing their trust to – what? Thin air? An alchemical breakthrough? Still, they may be right; sometimes we do need to throw the baby out with the bathwater and perhaps our vaunted sciences have done us more theoretical harm than good. Prudence, however, would seem to imply that we at least continue considering that science might have gotten some things right, and that we may learn more about the physical through it than by purely metaphysical speculations.

For these reasons I believe we can discount Seager's 'evidence' and by doing so strengthen our own view of consciousness as necessarily body bound (or, with a nod to agnosticism,

as anyway host bound), and thereby too we can tie the self in more tightly with the body. Nevertheless, wherever we place ourselves on this spectrum of the ignorant and the hopeful regarding what might become known, it does appear that we are able to (somewhat) confidently state that although we cannot definitively affix where or how consciousness comes to be it only does so relationally, at least as far as we know or can experience. This remains the case even if we grant dualism and say that consciousness supervenes on matter in some mysterious way: there is still simply no necessity for an intrinsic nature argument, though admittedly some may nevertheless find such an argument desirable. To think that consciousness in itself provides evidence of anything is to start with the assumption of a disembodied self, which Strawson of course does. Our self account has repeatedly argued for the impossibility of such disembodiment, at least in the human case, and if we are to accept human phenomenal data as evidence for a discussion of consciousness then surely we ought to maintain our focus on the human. To put it somewhat glibly, taken together what we have found so far is that standing *cogito ergo sum* on its head is what delivers a more accurate description of a human being: 'I am therefore I think.' To separate out and isolate the 'think', the consciousness, entirely from the 'I' leads only to the confusion that has so marked much of the philosophical treatment of this subject; Strawson's remark on long familiarity softening our critical approaches to a topic (made in reference to evolution and the experiential emerging from the non-experiential) seems apposite here. For historical and other reasons far too many thinkers may be starting from a presumed and unexamined Cartesian standpoint, to the detriment of all. This is another area where our discussion of consciousness is hamstrung from the first by the terms (and their background baggage) that we employ. I am not sure that an idea like Seager's 'relationalism' requires a full rejection of 'smallism' (he does state that consciousness' 'proprietary system of relations might well be a neural structure',³⁶⁸ which

³⁶⁸ Seager, *op. cit.*, p. 139.

seems to at least leave the door open for ‘smallest’ conceptions), nor that our only two options are it or some form of an intrinsic nature account; on this too I wish to remain uncommitted. For his part Seager finds that perhaps the main lesson that Strawson’s essay can teach us is that science cannot have all the answers; but then science does not really claim to.

Finally, on the note of what science actually says, the theoretical physicist H.P. Stapp responds to Strawson’s ideas through an explanation and application of modern quantum theory, starting with the charge that Strawson’s argument for his ‘real physicalism’ (what Strawson calls his panpsychist claims) as against standard ‘physicalism’ (empirical materialism) is based on what are now the outdated views of classical physics.³⁶⁹ This is important, I think, for although Strawson is a philosopher and not a physicist if he is to make the accusations against physics that he does (e.g. that we know nothing of the physical, that we have no evidence for the non-experiential, that physics has nothing to say on X, et cetera) then he really ought to be up to speed on the field, at least to the highest degree possible for a non-specialist. Even in his latter piece on panpsychism, discussed above and published ten years after Stapp raised the issue, the most recent resource on physics that Strawson cites is Stephen Hawking’s general audience book *A Brief History of Time*, published in 1988.³⁷⁰ That is of course not so very long ago, but in the world of contemporary advances in quantum physics (as well as the neurosciences) it is many multiple lifetimes. Stapp refers to the present consensus in his field as ‘orthodox ontologically construed quantum theory’, and relates that it does have within its lexicon and considerations experiences as they actually are or might occur in human beings’

³⁶⁹ H.P. Stapp, ‘Commentary on Strawson’s Target Article’, in *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?*, ed. by Anthony Freeman (Exeter, DEV: Imprint Academic, 2006), pp. 163-169.

³⁷⁰ Stephen Hawking, *A Brief History of Time* (New York: Bantam Books, 1988).

streams of consciousness, i.e. the choices of the experimenter/agent and the resultant feedback. The problem that many face in understanding it, Stapp writes, is the same problem that Papineau highlighted and which we considered above (and will again below): that of the terms being used.

Stapp summarizes this issue with, ‘What all this means is that the terms “physicalism”, “idealism” and “panpsychism” carry the huge baggage loaded upon them by centuries of essentially unquestioned acceptance of the basic concepts of classical physics’.³⁷¹ Such a vocabulary burdens us with inappropriate intuitions and conceptual frameworks upon which we base our judgments; it should therefore be no surprise that our judgments tend towards the erroneous. In an attempt to set things right, Stapp explains that the events of quantum theory are ‘experienced increments in knowledge’, and ‘the evolving quantum state represents *a state of knowledge*’ and also a set of tendencies (like potentialities) for new/future experiences. Stapp tells us that, ‘These considerations mean that a thinker with one viewpoint could call quantum theory a brand of *physicalism*, whereas a thinker with another viewpoint could call this very same theory a brand of *idealism*’.³⁷² As if the picture were not blurred enough, Stapp then goes on to write that Descartes’ idea of psychologically idea-like properties interfacing with mathematically described spacetime properties within the human brain is essentially accurate and quantum mechanics tells us how. It is just, I might add, very hard to imagine – but that might be the point, that our current imaginations, based as they are upon Newtonian models that work so well in the medium level world of scale where we live, are simply not up to the task. I, for one, can hardly follow Stapp’s piece and he very clearly simplified a great deal and included no advanced mathematics or even any mathematics at all (for which he must be thanked!).

³⁷¹ Stapp, *op. cit.*, p. 168.

³⁷² *ibid.*, pp. 167-168; emphases in the original.

The final note on Descartes' interacting model is very intriguing though, and we can conjecture from it that it might offer a glimpse on how the self may be real. In the concluding chapter of our study we will make a case for the self's reality that could conceivably be tied in here – depending again, possibly, on the thinker's point of view.

4.3 Seductive, but only a shimmering

On the whole then panpsychism seems to beckon us with a Siren's call. I will admit to personally being very attracted by the idea: not only do my phenomenal experiences seem to indicate that there is much more to the mental than the mere firings of neurons, and that surely mind is somehow more than brain, but also the notion that the universe is seeded with consciousness and even actually fundamentally composed of it is extremely satisfying on mystical and spiritual levels. As Agent Mulder's famous office poster in the television series *The X-Files*³⁷³ put it: I want to believe. Unfortunately the idea does not seem tenable unless we make what I consider to be too many initial assumptions and dismiss what I consider to be too much of contemporary empirically grounded evidence. Most problematically, at least within the confines of the present study, is the inevitable result that to accept panpsychism seems to lead to accepting mind as an existent separate from and without requiring a body, and if there is mind then we must also look for subjecthood and self. For reasons already given in Chapter 3 I do not find Strawson's thin self/minimal subject account – which would answer the subjecthood question – convincing, and although I would be comforted by the concept of the universe as a single and self-contained being (universal self of which all selves are parts or facets), I am not sure that such a belief is warranted. If it is then we would have the subject needed for an all-embracing and all-constituting consciousness; on that too it is probably best to remain

³⁷³ For a brief summary of the show see 'The X-Files', *Wikipedia: The Free Encyclopedia*. <https://en.wikipedia.org/wiki/The_X-Files>.

open for now. Conversely, the tendency towards the development of consciousness at particular levels of organic complexity might be what is 'built into' the universe, and if so then it would seem that emergence would be a good explanation for how. Perhaps quantum mechanics will provide some assistance on the point and perhaps in time we will have arrived at sufficiently altered ways of viewing the physical world that such answers will make sense (intuitively and otherwise) to those of us who remain non-specialists.

The question of the nature of consciousness is almost as perplexing as that of its origin, and despite all of our panpsychist considerations we do not seem any closer to really knowing anything, although we have at least learned what we do not know. The next chapter will therefore survey what contemporary neuroscience can tell us about consciousness, and then apply what we have gleaned to consciousness' experiential side. Throughout we will maintain our attention on the human (except for some brief comments when working on definitions derived from neuroscience), on what consciousness is for us as the creatures that we are. To do otherwise, I think, would be to risk theoretical overextension, particularly given that the empirical studies we have examined and taken into account have also been focused on the human case. Any attempt at cross-species universalization could be just as likely to lead to error as an incautious use of terminology. Consciousness and its relation to the self is after all a difficult enough topic, we need not make the terrain any more foreboding.

Chapter 5: Consciousness, Qualia, and the Self

5.1 Modern neuroscience and the structure of consciousness

A. Consciousness, intent, thinking: Definitions

Although the explanatory picture of consciousness now given by neuroscience is quite naturally a work in progress, as remarked on earlier an increasingly consensual view has developed in recent years, and in many ways it is quite illuminating regarding the root biology involved. We have previously glanced at this picture, and our exploration of panpsychism and what place consciousness might and might not have constitutively in the natural order has added much to the modular view that was offered. Yet there is still much ground to cover if we are to come to any conclusions on the matter – however tentative – and so in this chapter we will now spend much more time looking directly at this cognitive portrait, gazing at its details, taking in its brushstrokes, studying its layers and nuances, all with the goal of uncovering this part of the self that, on our definition at least, is fundamental. Many philosophers (of mind but also in other areas) have sadly eschewed this aspect of the human animal, either remaining largely uninformed of the science involved or rejecting it or its application based on the judgment that biological expositions cannot account for experience. If that is so, if they cannot, it is thought, there is little to be gained from them. This is unfortunate, and as we will discover, very far from being accurate. Let us therefore take up our place before this image and begin our study of it, starting with the biological model and related necessary definitions.

The (in)famous purported failure of scientific accounts to handle experience has been termed the ‘gap’ or ‘hard problem’ of consciousness, and it will be discussed in the below, but prior to that I wish to again highlight that one of the main, and perhaps *the* main,

difficulties that we confront when we look at consciousness is that we have an unhelpful tendency to confuse ourselves right at the outset by the manner in which we talk about the issues under consideration. What I mean is that I think we tend to set out from, or to overly focus on, the wrong level (in one of two ways), and in so doing we present ourselves with a seemingly unsolvable mystery or a seemingly incomprehensible simplification. I will get to each of these levels in turn, but to do so in a manner that makes the most sense (to me anyway) I initially want to lay out what the terms commonly used in discussions of consciousness indicate on my own analysis of the issue since I think that it is from these terms, or rather from the inconsistent employment of them, that much of the uncertainty arises. If we can get clear, or clearer, on consciousness then our resulting perspective on the self will be much aided and our self theory greatly informed. First and foremost, I do not find consciousness to be thinking: thinking is rather one potentiality of organisms with the ability to manipulate the information that their brain receives and in turn produces; consciousness is more fundamental, as the outline of it and the place we have given it thus far in our study would indicate. However that statement just in itself is not likely to be very transparent, and so let us take a moment to look at how the brain processes information.

All organisms with senses of some kind, no matter how rudimentary (a single-celled organism moving away from danger or towards food), receive input that is processed by or in a region specializing in that processing. A cat smells a mouse and her brain triggers a certain response. A cockroach's antennae alert him to movement in the vicinity and his brain triggers a certain response. A bat hears a particular pinging and her brain triggers a certain response. A paramecium's cilia sense an acid nearby and its 'brain' triggers a certain response. The reader will have noted that the final 'brain' is in quotes/inverted

commas,³⁷⁴ and the reader will probably not have found anything amiss with that – yet why not also do so for the cockroach? Can his brain really compare to that of a cat’s or a bat’s? Of course it cannot, but it is a very far cry from the simplicity of a paramecium’s two nuclei, and that much is easily recognizable and just as easily agreeable. Reasoning about why that might be the case though brings doubts to our minds, especially when we consider that the paramecium’s nuclei act in much the same information processing and directing manner that a more acceptable ‘brain’ does, the difference being one of degree, and that of degree of complexity. To get through this potential logjam let us establish that a brain – as reflections of this sort lead us to conclude – is an information processing tool that handles the input it receives and then generates, for better or for worse, a responding output. A sense interacts with something in its external environment, feeds that interaction via chemical channels into itself (body), a brain works on what has been transmitted to it, sends out from itself another chemical signal, and other parts of the body react in some manner (observable or not). This is how life works, and it is easy to see why life should work this way because all life happens in an environment of some sort that is external to the life inhabiting it. The very same life also has an internal environment though, and the very same brain (of whatever sort) is also always receiving input from it – volume and complication are beside the point, each brain processing unit is getting signals from outside its body and from inside its body all the time. In fact, nearly all of those inputs are from the inside, for once past the initial sensing or perceiving organ/boundary all that commences is internally accomplished.³⁷⁵ Biology 101; yet we forget these important points and jump

³⁷⁴ The reader may also have noted the use of gendered language for all but the paramecium; given the presence of genders in these organisms I find the use of such appropriate (and indeed, instructive of how we might wish to think of our place in the world, but that is another topic).

³⁷⁵ Douglas Hofstadter emphasizes this point in his *I Am A Strange Loop* (New York: Basic Books, 2007). On the neurological picture and some related issues described here and below, in addition to Hofstadter’s work see Gazzaniga, *op. cit.*; V.S. Ramachandran, *The Tell-Tale Brain: Unlocking the Mystery of Human Nature* (London: Windmill Books, 2011); António Damásio, *Self Comes to Mind: Constructing the Conscious Brain* (New York: Vintage Books, 2012); Strawson et al., 2006, *op. cit.*; Dennett, *op. cit.*; and David J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory* (Oxford: Oxford University Press, 1996).

right to how we move and – crucially – *feel* in our world, and this is where I think we begin to trip ourselves up.

We have these senses and they are sending information to our central processing units, all well and good, but what about consciousness? The fuller model I wish now to present (with all of my limitations as a barely informed philosopher and a non-specialist), the model as conceived after our many considerations thus far and in what follows, is a many layered one. Underpinning this view is how the brain itself deals with the information it receives, and that is in the production of representative maps by the flexible and shifting use of neurons.

A neuron is a very special kind of cell, not found in all living creatures and not essential to life processes, but they have come to function in complicated creatures as assistants in the management of the many other necessary cells that make up the organism. Groups of neurons can be molded (switched on or off is the usual terminology) in countless ways to help the brain process now this, now that. An incoming signal will cause the relevant area of a brain to generate within its causal pathways certain chains of events that (could) result in X neurons being grouped (or ‘switched on’) in such a manner that symbolically represents the pertinent information as a map upon which the brain can operate in its processing and determination of output. Though such might primarily be handled by a specialized region in the brain, the brain’s pathways importantly cross and intersect its many districts, with all areas linked and in constant communication via internal signaling. The brain as a whole is, after all, a manager, and as a manager it is built to respond, and like any good manager it needs to be able to handle vast amounts of information adequately when it makes its responses; the biological key to all this, as indicated, is the

formation and utilization of (neuron constructed) representative maps.³⁷⁶ Consciousness, I think, is to be found at this level, but just ‘where’ or ‘how’ will become clearer in the section below on the so-called gap between the level of neurons and that of experience (I do not think either ‘where’ or ‘how’ are appropriate questions but they are the natural responses); for the moment I only want to posit that the brain’s networked use of representational tools – tools that are constantly being formed and re-formed for the purpose of directing the organism in its environment – is the foundation of consciousness.

Note that this means that not all living organisms have consciousness, and certainly that not all of the universe is conscious, at least not in the way we know consciousness. It also, however, means that just as some organisms’ map-making abilities are more rudimentary than ours (but still there), consciousness appears to come in different degrees, for if the production and utilization of representative maps is what anchors (emergent) consciousness then a dog, say, with her simpler maps, has a less sophisticated consciousness than a chimpanzee’s, and we might postulate that our maps could be a dog’s version compared to another organism’s – the universe is a big place, after all. What is significant is the fundamental structural level of consciousness in creatures who have it, to whatever extent. ‘Consciousness’ is therefore the label we use to describe the procedural application of these tools by the brain functioning in its interlaced entirety. This may be why emergence appears to be such a good explanation for where consciousness ‘comes from’: rather than asserting some kind of fantastical process the term instead signifies that we have moved in thought from a biological point of view to a representational point of view – that is to say, we have only changed how we are looking at the brain (parts/whole), we have not hoisted something onto it.

³⁷⁶ Damásio *ibid.* is quite clear in his description of maps and their roles, as he is in explaining the specialized structure and purpose of neurons.

We are not yet at this point (or layer) discussing intent, and that is a meaningful detail as intent, consciousness, will, thought, are all too often conflated in our thinking and our discussions. Intent I place a step up from consciousness, and root it in the organism's push towards thriving, to the pursuit of its (unique) pleasure and the avoidance of its (typically less unique) pain. Some thinkers attribute a kind of 'natural will' to all life that arises from the urge to survive and to reproduce one's kind,³⁷⁷ but I find such to be almost a form of anthropomorphism for to say that a paramecium actively 'wants' anything is, on reflection, to stretch the concept of 'want' beyond the breaking point given the term's implications of connected thought and planning. We might instead speak of a 'drive' but that is hardly much better, and given that just as we cannot imagine what it would be like to be a paramecium and operate without any of the mental capacities that we take for granted, and that a paramecium moreover cannot itself think nor imagine anything at all, it is probably safest and wisest to just leave it at the juncture where we simply recognize that life is built to keep going. Intent therefore is the motivated behavioral purpose or design that comes from an organism's constant body monitoring when it has a functioning brain, a working central processor and manager, and it naturally differs widely across time for each kind of organism and for each individual within that kind. The cat now wishes to lick her fur while sunbathing, the cockroach to return to his nest, the bat to obtain a better perch, and those wishes are based on the brain's analysis of the current overall state of affairs and used as directing impulses/commands towards engaging in the according output. Intent needs consciousness, it is not consciousness.

Finally, then, thinking. Just as consciousness is neither universally found nor universally equal, so too for thinking. The incessant body monitoring that the brain engages in allows

³⁷⁷ See, for example, Damásio *ibid.*, especially Chapter 2 ('From Life Regulation to Biological Value').

for degrees of reflection on one's own being in those creatures whose brains make use of representative maps, though understanding the extent to which such self-reflection is possible for other species may always remain out of our reach. Does a dog know that he is happy and therefore wags his tail as an expression of that, or does he rather automatically wag his tail because he is happy? Whichever is the case for dogs, for us we find ourselves in the position of not only being able to consider what is mentally represented but to consider representations of representations, and it is the considerations born out of this symbolic and abstract processing ability that we call 'thinking'. Our brains are capable of a limitless degree of representation,³⁷⁸ and although that representation is rooted (evolutionarily) in internal biological monitoring vis-à-vis the environment and organism health, the skill itself is hardly limited by its originary purpose. In sum, consciousness is not thinking, to equate the two is to make the same mistake as that which limits consciousness only to mental acts like a System 2 analysis: we have labeled this level 'awareness' and taken pains to show how consciousness at large runs much deeper and broader. Still, thinking could not exist without consciousness, and – as everyday occurrences indicate – thinking comes in many forms and in many versions. With that the specter of experience has once again been raised, and so let us turn to the idea (thinking) of the so-called gap, which is perhaps the real crux of the matter for such a plethora of writers working on issues related to consciousness.

B. The gap, the hard problem, the neuronal to experience straits: Navigating

The gap in question, as mentioned, is that between the world of neurons and biological brain functioning and that of our felt experiences as we go about the business of living. At its core this is the question, really, of qualia. We cannot imagine how it could be that an accumulation of sticky brain bits and pieces, which taken cell by cell have neither

³⁷⁸ On this see especially Hofstadter, *op. cit.*

consciousness nor experience, could work together in such a way that a rose when viewed gives its beholder a subtle feeling of comfort and appreciation of beauty, that a chocolate does not merely taste sweet but is also soothing, that a photograph evokes a memory of a time and a place that fills us with nostalgia. All of these things are descriptions of the ‘what it is like’ aspect of qualia,³⁷⁹ and they are necessarily internally bound and contingent upon the experiencing individual – my interaction with a rose, though perhaps quite similar in recognizable ways, will differ from yours (though ours will be far more like each other’s than like a cow’s involvement with the same rose; what I am attempting to get at is the personal side to qualia based on each individual’s unique particulars). Clearly all of these concerns are moreover tightly connected to the self, particularly on our view with its three informing and influential emotive sets, and I believe we will be able to find that an investigation into this area will yield some answers to our queries on the self-constituting makeup of consciousness and on its relation to the other formative elements of the self.

Yet merely knowing intellectually that our brains make use of neurons to create representative maps that provide them with usefully abstracted information upon which they can function and thereby better manage our beings in our environments does nothing to help overcome what seems for all the world to be a tremendous chasm between that cellular level and our enacted lives. What does any of it really practically mean? We have no analogue to what is going on inside our brains and so we cannot properly think about it,³⁸⁰ we cannot pull our brains apart as we might an engine to see how it all works because there are far too many parts involved and the details of the interplay between those parts remain far too opaque. This is the central baffling mystery, and it is a very old one, though

³⁷⁹ This concept was first brought to wide attention by Thomas Nagel in his wonderful 1974 essay ‘What is it like to be a bat?’, collected in *Mortal Questions* (New York: Cambridge University Press, 1979), pp. 165-180.

³⁸⁰ Hofstadter, *op. cit.*

its contemporary groundwork (setting the direction for the previous century and the present one) was perhaps best laid in the questioning of the physical raised by luminaries like Arthur Eddington and Bertrand Russell.³⁸¹ It has led even giants of otherwise highly physico-rational thought such as Nagel to proclaim that our physical picture must be wrong, that what we know cannot possibly be all there is to it, and that it should therefore be questioned and perhaps (even entirely) reworked.³⁸² An excerpt of an interview by Tim Parks of Riccardo Manzotti that deals with this issue via the experience of color and that appeared on the *New York Review of Books* website gives a good example of just how unfathomable this gap is taken to be:

Parks: I appreciate that you've spent a great deal of time researching the history of science's dealings with color, but are you telling me that contemporary neuroscience offers no dominant view on the matter?

Manzotti: Well, the current textbook view goes like this. The world is a place where objects reflect light, sunlight being the dominant source and as it were default setting as far as the kind of light is concerned. However, each object reflects only a subset of that light. Rays from this subset enter our retina and stimulate a honeycomb of cells, known as cones, because of their conical shape, whose function is to react differently to different portions of the visible spectrum (we remember of course that only a small part of the vast electromagnetic spectrum is visible). Most humans—animals are rather different of course—possess three kinds of cones, referred to as S, M, and L cones, depending on whether they react more vigorously to short, medium or long-range light wavelengths. The “output” of these cells is first merged together in the retina then sent

³⁸¹ Arthur Eddington, *The Nature of the Physical World* (New York: Macmillan, 1928); Bertrand Russell, *The Analysis of Matter* (London: Routledge, 1927/1992).

³⁸² This call was made in Nagel 2012, *op. cit.*

via the optical nerve to various cortical areas—including the famous V4. And that’s as much as we know.

Parks: Riccardo, you just gave me the whole explanation without ever using the word color.

Manzotti: I know! Oddly, this is a theory of color that does not need the notion of colors. I suppose the reason is that however carefully you follow neural signals from the retina along the optic nerve and across the brain you don’t actually come across anything like a color, or anything that explains color perception. You could almost say that the notion of color is useless to color science, unless...

Parks: Unless?

Manzotti: Unless we bring consciousness back in the picture. Colors are something we *experience*, individually and collectively. But without our experience of color, science would have no reason to suspect its existence. There would just be fifty shades, or more likely fifty thousand shades, of electromagnetic waves.³⁸³

To my mind, the purported oddity that is assigned to the given description of how light is processed by one’s eyes and brain arises from what appears to be an approach to the issue by both Parks and Manzotti from the higher, abstract level of the concept ‘color’ and not from the cellular level on which the processing itself takes place. They seem to be thinking,

³⁸³ Riccardo Manzotti and Tim Parks, ‘The Color of Consciousness’, *The New York Review of Books: NYR Daily*, 08 December 2016.
<http://www.nybooks.com/daily/2016/12/08/color-of-consciousness/?utm_content=buffer03030&utm_medium=social&utm_source=facebook.com&utm_campaign=buffer.>; emphasis in the original.

and hence are talking, on a level that is not apposite to their topic: that of the end product experience of color perception rather than its biological and informational input aspect. They are apparently looking at the output and asking how it could possibly have been made from the interaction of the sense with the stimulus and its physical processing. They are alleging – or implying – that the brain, as a mere physical thing, cannot be connected with consciousness since consciousness is not itself physical. They are evidently seeing an ‘extra’ that is claimed to proceed from the physical, they are asserting that in all of our described experiences the physical picture can do little better than to stubbornly insist, without explanation, on a something coming out of a nothing.³⁸⁴ The closeness of what is being assumed here with emergent explanations for other phenomena should be noted but not, I think, overly emphasized at this point. It is my hope that in the below all or many of the facets we have examined so far will come together into (at least some manner of) a coherent theoretical perspective.

The tendency to approach the alleged problem of experience from the point of view of the abstract and symbolic level illustrated by the above-quoted portion is where I think our troubles come from, and it is an entirely understandable tendency since it is on that level that we operate in our daily lives given our brains’ map-making representational procedures. As the creatures we are we function solely on this higher level of tokens and their applications as we negotiate our physical and social (and mental) environments, meaning that we are oblivious to the physico-chemical in our thought and in our speech. Douglas Hofstadter, in making this point, puts it this way:

This, our innate blindness to the world of the tiny, forces us to hallucinate a profound

³⁸⁴ See the whole interview, and the one that preceded it in their series, for full details. The first installment is called, ‘The Challenge of Consciousness’, and can be found here: <http://www.nybooks.com/daily/2016/11/21/challenge-of-defining-consciousness/>.

schism between the goal-lacking material world of balls and sticks and sounds and light, on the one hand, and a goal-pervaded abstract world of hopes and beliefs and joys and fears, on the other, in which radically different sorts of causality seem to reign.³⁸⁵

The key to bridging this schism, to seeing that there is really not one there at all, is to recognize that there are not in fact any ‘radically different sorts of causality’ present (to think that there are is to look from the wrong level, or to fail to shift levels between the default conceptual and the more abstruse biological); yet to get to that point we must first understand the depth of emotional processing in the brain and what that in turn entails for our experiences – our feelings of ‘what it is like’ – as we become aware of them in self-reflection. The going may not be easy, but will, with our background, be doable.

As outlined in previous chapters (here briefly recapped), recent psychological research into decision-making and the root sources of behavior has revealed that the brain operates in its relation to the world in two different ways: one is the automatic, efficient, and preconscious (i.e. pre-aware, or unaware) method that is dominant and is often referred to as System 1, and the other is the regulated, slow, and laborious method of rational analysis, often called System 2. Or, in other words, the intuitive and emotional method and the reasoning-based method.³⁸⁶ What is most crucial for our present purposes are two points in relation to this overall picture, namely 1) that due to the real-time pressures we have faced in the course of our evolution System 1, with its rapid judgments and responses that bypass the need to stop and think, is by far the more frequent decision-making apparatus

³⁸⁵ Hofstadter, *op. cit.*, p. 204.

³⁸⁶ The psychological research in this area has become quite dense and many references have already been given; some of the primary sources that have informed my thought on this recurring point are: Haidt 2001 and 2012, *op. cit.*; Damásio 1994 and 1999, *op. cit.*; Tversky and Kahneman, *op. cit.*; Kahneman 2003 and 2011, *op. cit.*; Dijksterhuis, *op. cit.*; Gary A. Klein, ‘A Recognition-Primed Decision (RPD) Model of Rapid Decision Making’, in *Decision Making in Action: Models and Methods*, ed. by Gary Klein, Judith Orasanu and Roberta Calderwood (New York: Ablex Publishing Corp., 1993), 138-147, see also his Klein 1998, *op. cit.*; and Greene, *op. cit.*

upon which we operate (indeed many nonhuman animals who lack the means for System 2 reasoning spend their whole lives quite successfully engaged solely on it), and 2) that due to its preponderance the unceasing results of System 1 analyses conducted by the brain as it processes the information it receives are always included as one of the evaluative elements if/when a later System 2 analysis is conducted. What this means in practice is that for every stimuli/perception encountered the brain very quickly conducts an entirely preconscious analysis based on genetically equipped biases, information from previous encounters, relevant knowledge/memories, et cetera, and delivers up a judgment, an intuition, a ‘gut feeling’ regarding the stimulus/percept in question; this could be something very simple (fight/flight) or more complex (approach and investigate with/without caution), but it will always be there. Nearly all of the time this is how we operate: automatically and non-rationally.

As was discussed, this process has important implications for one’s self-view, for one’s behavior, and for the full self (the self, personal identity, whole person) embedded in the world. Our brains, however, also have their System 2 components, and as described earlier, System 2 is capable of overriding the judgments of System 1, but its analyses will still include within them the results of the preceding and unavoidable System 1 judgments. Again, in practical terms we might encounter a strange-looking object on the sidewalk on the way to catch a train and be curious as to what it is, so intrigued, in fact, that as we approach it we begin to think whether or not we have time to pause and look at it, maybe even to handle it, maybe even to pick it up and keep it. In our considerations on this point a number of factors will play their part, but amongst them will be the initial result of whatever our System 1 analysis determined, and we will experience that result as an intuition – not as a known but rather as a felt inclination in one direction or another. That feeling will influence whatever rational procedures we bring to bear on the decision. This

is the emotional face of the brain, and it is constantly present: now as a general and almost aethereal background mood, now as an overwhelming compulsion. This input too, as we have seen, is also importantly informative of the self's constitutional elements. When we therefore regard the brain in the dichotic terms of common parlance we find that it is far more of a tool for feeling than it is a tool for thinking, though in truth the line between the two (feeling and thinking) is far blurrier than is usually supposed, as I hope the preceding has shown. Herein lies one answer to the presumed quandary of qualia.

Experience, on our multi-layered model, operates at a step down from thinking, it is the pre-reflective and automatic (physico-chemical) reaction to the received input that the brain has processed; it is far less abstract than thought, though it is still based on the brain's system of representative maps. It is also far more directly connected to the body proper – the physical organism in its environment – as it is the result of automatic analyses done in the interests of life management. The reader will recall that even when dealing with external perceptions the process is almost entirely an internal one (once sensed the signal begins its fully internal journey to the brain, its fully internal processing by the brain, and its fully internal response signal sent back by the brain); and the brain manages the body by use of the intuitive judgments and emotional tags that it assigns to absolutely everything. A strong feeling of aversion is a far more efficient, and far more effective, means of avoidance than is taking the time to stop what one is doing and think about the merits and demerits involved before arriving at a costs/benefits based decision and only then taking action (or concluding not to take action). Everything feels as it does to us because the use of feelings and other automatic affective distortions are one primary way that the brain does its job of directing the body, and since the signals sent internally to the brain and the brain's internally sent signals in response are so remarkably intertwined within ourselves

every percept, every stimulus, every experience, begins to feel *like* something.³⁸⁷ Qualia. More specifically, the physical basis for qualia; more generally, the disappearance of the chasm, the crossing of the schism, the closing of the gap. These feelings are biologically caused and biologically expressed, and it is only in our descriptively related experiences of them – the point at which we bring in the conceptual level, the symbolic and abstract – that we begin to consider them to be otherwise. Moreover, while you might have differing personal associations with roses than I do due to your individual past (perhaps a lover jilted you by curtly returning the dozen roses you sent him), and hence will experience them in an altered way from me, both of us will have a physico-chemical response to roses and our brains will process such with the attending intuitions and emotions applied. Once again, the process is a biological one and the qualia generated are likewise biological, even if we describe them in terms that are very much non-biological. There is thus no difference between what a feeling *is* and what a feeling *feels like*; there is only a difference in how we think about and discuss them.

C. An aside: Always conscious?

I have taken the position in this study that if at least one of the many specialized brain regions (modules) that make up the human ‘constellation’ of consciousness (using Gazzaniga’s evocative metaphor) is working then we ought to consider ourselves to have consciousness, and since each area of the brain is fully dormant simultaneously only in the case of death the result reached was that we ought to regard ourselves to *always* have consciousness, even if we are not always ‘conscious’ (i.e. aware) at all times. The seemingly obvious rejoinder to such a position is the experience of sleep or a blackout or another common occurrence in which it is thought that one clearly does not have consciousness; the reference in such an objection, however, is to ‘consciousness’ in its

³⁸⁷ Damásio 2012, *op. cit.*

unanalyzed common usage meaning of awareness, not in its fundamental structural position as outlined above.

What is interesting to note, moreover, is that one of the same brain regions associated with awareness is active during dreams. Damásio lists three core areas of the brain as being consciousness-making: the PMCs (posteromedial cortices), the thalamus, and the brain stem; in the dream state, he writes, our brains apparently make active use of the PMCs.³⁸⁸ Marcus Raichle and Debra Gusnard's research also points to a 'default network' within the brain that is constantly active and which, perhaps counterintuitively, actually becomes slightly depressed when one's attention is specifically directed.³⁸⁹ Damásio proposes that the PMCs are an 'important integrator/coordinator' for consciousness that 'would remain active at all times, attempting to hold highly disparate sets of background activity in a coherent pattern.'³⁹⁰

Here again it seems that an imprecise, or inconsistent, use of terminology when it comes to consciousness can yield conflicting viewpoints. If 'consciousness' is used in reference to a structural state of the brain's functioning (foundational within the layered model presented above) then the evidence does appear to point to our 'always being conscious' as a correct description; if, on the other hand, 'consciousness' is used synonymously with 'awareness' (as in the phenomenological 'conscious of') then the description would not be apt as quite naturally we are not always aware of our surroundings and/or similar other elements. The point might not seem immediately pertinent, but it is indeed directly related to our concerns as I will later give an account of how our self-view can be considered realist, and

³⁸⁸ Damásio 2012, *op. cit.*; note that this is presented as his view, I am unsure if/to what extent a consensus exists amongst neuroscientists on this detail.

³⁸⁹ D.A. Gusnard and M. E. Raichle, 'Searching for a Baseline: Functional Imaging and the Resting Human Brain', *Nature Reviews Neuroscience*, 2 (2001), 685-694.

³⁹⁰ Damásio 2012, *op. cit.*, p. 243.

since one of the definitional elements of our self concept is consciousness (Set 1 + Sets 2 & 3 + C + B_p = the self) it is in fact critical for our substantive self that consciousness be present, else the self might not be – and then hence not real.

In reflecting on that though we can already see that we may run into other trouble as the three emotive sets which are no less definitional could conceivably be absent in a state like anesthesia or even sleep, but that discussion can be saved for our final chapter. The concluding part of this chapter will instead explore qualia from a more standardly philosophical perspective, examining the significance of the physical basis for qualia as found above and the centrality of qualia to the self, and then lastly begin to wonder about how the self might be real.

5.2 Qualia and the self

A. ‘What it is like’

As mentioned above, Nagel’s very famous ‘What is it like to be a bat?’ paper³⁹¹ brought philosophers’ attention to the question of ‘what it is like’, which is the question of qualia. That paper, having the foundational position that it does, will accordingly be centrally considered in the below as we shift from a philosophically-informed neurological perspective on qualia to a neurologically-informed philosophical perspective. This section will therefore consist of some thoughts on Nagel’s piece, some representational objections that have been made to it, and some comments on how those objections may have missed the point. What will be attempted will be a greater degree of clarity regarding the phenomenology of consciousness, of what it is like for a human being to experience consciousness understood in the sense of awareness that we have been using.³⁹² Having,

³⁹¹ Nagel 1979, ‘What is it like to be a bat?’, *op. cit.*

³⁹² We will continue to employ our more restricted awareness based definition of ‘consciousness’ here since consciousness’ root or core functioning (the underlying layer we have presented) does not always enter into

hopefully, achieved that clarity we will then attempt to tie in those conclusions with our broader goal of defining the self via reflecting on the contemporary question of artificial intelligence and the often-raised analogy of the mind-body relationship with that between software and hardware.

The term ‘qualia’ has been defined in the philosophical literature as an abbreviation of ‘phenomenal qualities’, and refers to the specific sense that is associated internally with a mental state; as David Chalmers puts it, ‘a *qualitative feel* – an associated quality of experience... The problem of explaining these phenomenal qualities is just the problem of explaining consciousness.’³⁹³ In this Chalmers conceptually equates the having of qualia with the having of consciousness, and this agrees with Nagel’s own view as expressed in his ‘what it is like’ notion, defined by him as follows: ‘But fundamentally an organism has conscious mental states if and only if there is something that it is like to *be* that organism – something it is like *for* the organism.’³⁹⁴ Notice the intuitive appeal of this approach: it does indeed seem to us that our mental states carry certain ‘feels’ to them, that there is something it is like for us to be happy, to suddenly remember a task or event, to notice a dropped pen, et cetera. Everything does (when viewed from the conceptual level) ‘feel like’ something. Nagel, Chalmers, and others then associate the subjective – the specially subjective, the only subjective, the unshareable – nature of such experiences with *consciousness itself*, consciousness just is ‘what it is like’ (that is, in the hard problem sense). Whether or not this definitional move is warranted is a question we will consider in the following objections to Nagel’s position, however prior to that I would like to first offer an alternative position on what Nagel may have meant for I think that he, and this piece of his in particular, calls for an understanding framed otherwise from the standard one.

descriptive/describable experiences, and if it does not there would be no ‘what it is like’ about it.

³⁹³ Chalmers, *op. cit.*, p. 4.

³⁹⁴ Nagel 1979, ‘What is it like to be a bat?’, *op. cit.*, p. 166.

The idea of qualia has been taken to mean, or has been asserted to mean, that in addition to the obvious experiential qualities of a phenomenon like feeling happy or being surprised, specific thoughts have their own associated, and again singular, content. Going back to Chalmers we find, ‘When I think of a lion, for instance, there seems to be a whiff of leonine quality to my phenomenology: what it is like to think of a lion is subtly different from what it is like to think of the Eiffel tower.’³⁹⁵ My own view is that this level of particularity is not the kind of thing that we should take Nagel to have in mind in his ‘bat’ piece, and in the following I will attempt to show why.

In the article itself Nagel, in the midst of writing about a bat’s perception, pain, fear, hunger, lust, states that each have a certain internally felt *thusness* to them (in his words: ‘we believe that these experiences also have in each case a specific subjective character’), and he follows this with a comparison to a hearing and seeing person with ‘the experience of a person deaf and blind from birth’ and the inaccessibility of cross experienceability for each, although each knows that the other has subjective experiences.³⁹⁶ While this latter comparison may at first blush appear to make it seem as if Nagel is offering the same sort of claim that Chalmers does a reminder of the context will help us see that he may not be, and that conclusion results in some important implications for the concept of qualia itself.

Prior to this comparison Nagel’s examples have all been at a much higher and more generalized level than that of a thought of a lion compared with a thought of the Eiffel tower; we are dealing – in Nagel’s piece at least – with how a bat perceives via its sonar, how a seeing person experiences the world through vision, how a non-seeing person

³⁹⁵ Chalmers, *op. cit.*, p. 10.

³⁹⁶ Nagel 1979, ‘What is it like to be a bat?’, *op. cit.*, pp. 169-170.

experiences the world through sound and touch. Does that mean that a seeing person and a non-seeing person cannot know what chocolate tastes like to the other? Of course not, for both have human tongues located within human bodies and both know that the other's acknowledgedly subjective experience will therefore be similar to their own as the same functioning organ is in play for each. Perhaps there will be slight variations between the two (as with our rose earlier), but that has nothing to do with the visually, aurally, and tactilely perceptual where their primary differences lie.³⁹⁷ This is unlike a person and a bat where the variances extend far beyond any perceptual issues alone. We must note that Nagel does not say that *each experience of* pain, fear, et cetera has a 'specific subjective character'; the 'each case' in his usage seems instead to refer to categorical *type* (pain, fear, et cetera generally; and in that surely pain for us as humans, pain for bats as bats). On my understanding of it this is not a splicing of *this pain* versus *that pain*; it is instead *of pain, of fear, of hunger*, and on and on. This 'what it is like' is a statement of a broad phenomenology, experienced subjectively amongst individuals but set at, and focused on, the species level as Nagel's examples of intelligent bats and Martians go on to show (discussed shortly).

I clearly remember the moment when I had this insight – if such it is –, and the experience of it was not a qualia of realizing something about Nagel while reading Nagel, nor even of realizing something about a philosophical point while reading a philosophical text; it was, rather, the same sensation that I have had before (sadly far too infrequently) when my thoughts have clicked over afresh, opened an unseen door, looked through a hitherto opaque wall, that is associated with a certain sense of awe and a feeling of time coming to a standstill (more on that very important feeling below). To reduce the notion of qualia

³⁹⁷ Or anyway it has very little to do with it; the sight and/or smell and/or touch of chocolate may generate a certain preparatory excitement in one or the other.

beyond the idea of a type is to go further with the sense of ‘what it is like’ than I believe Nagel would and, as will hopefully become apparent, than is warranted given what is actually going on in our phenomenal experiences.

This reading of type rather than the more standard ‘whiff of leonine’ exegesis is supported, I believe, in what immediately follows in Nagel’s original text. There, within the same context and indeed in the very next paragraph, Nagel comments on the ‘certain general types of mental state [that] could be ascribed to us’ by intelligent bats or Martians comparing themselves with us and only granting us the types they saw or could see in both: e.g. ‘perhaps perception and appetite.’³⁹⁸ Again, the argumentative focus is on a general level, not minutiae. While it is admittedly possible that in this Nagel meant that each particular mental state has its own, and very peculiar, feel to it but that such would not be noticed by beings as different from us as intelligent bats or Martians, given the examples of experience used (and cited above: perception, pain, fear, hunger, lust, appetite) that are said to have ‘specific subjective character’, together with his thoughts on what might be ascribed to us by very different beings engaged in observing us, I believe that although it is not entirely clear the case for my understanding of Nagel might be the stronger one given the considerations outlined. It is at least another way of reading Nagel. If my interpretation is valid then qualia ought to remain on the level of type (which is phenomenally wider), and not be extended down to the microscopic degree they have been. A (re-)understanding of qualia in this way would engender significant consequences.

I mentioned a certain sense of awe and a feeling of time coming to a standstill that was associated with my alleged insight into Nagel’s text; such concurrent affective experiences are paramount in what they teach us about qualia and the implications that they have for

³⁹⁸ *ibid.*, p. 170.

the concept as the very notion of qualia has been challenged on the same phenomenal grounds. (It might be thought that I was making a similar critique above when discussing what a feeling *is* and *feels like*; I was not, as will be expounded below, but the point made there about the lack of a difference is nevertheless a crucial one.) Before considering some representative objections to Nagel's paper and what has been taken from it though, I would like to pose a question by way of recap of the above: If the quale of a thought differs depending on the subject of the thought, is that difference not due to the associated emotional content rather than the thought content? As I see it, if it is like anything to think then the act of thinking itself is *like* something, not thinking of this as opposed to thinking of that (hence qualia as type). It might be that there are special times when a thought has a more profound emotional impact than is typical – such as the experience I described –, or that I have a certain emotional juxtaposition with a thought about, say, an object that you do not have, but on the whole thinking has its feel, perceiving its, pain its, fear its, et cetera. On my understanding there should be no more additionally detailed reduction of the notion of qualia beyond that point, beyond that more generalized level of type. There are some who go further than this in their challenge to the concept though, and so let us now take a look at some criticisms and attempt to answer them.

P.M.S. Hacker disagrees broadly with the idea of qualia, arguing that the difference between, say, seeing a table and seeing a chair does not consist in a different sort of associated feeling as just the bare perception typically fails to bring about any sort of emotional or attitudinal reaction in us.³⁹⁹ Perception itself, for him, does not equal having a sensation. Moreover, he continues, different experiences may have the same feel, being as 'enjoyable or disagreeable, interesting or boring'⁴⁰⁰ so that we suspect it is not the feel

³⁹⁹ P.M.S. Hacker, 'Is there anything it is like to be a bat?', *Philosophy*, 77:300 (2002): 157-174.

⁴⁰⁰ *ibid.*, p. 164.

itself, when there is such a feel, that sets them apart either. These are good observations and should be taken seriously. Hacker goes on to argue that when speaking of ‘what it is like’ we cannot say, ‘(1) “There is something which it is like to V”’ and, ‘(2) “There is something it is like for A to V”’ because in the case of (1) such statements are fit only for comparisons, and in the case of (2) such statements wrongly mix, ‘the form of a judgment of similarity with the form of a request for an affective attitudinal characterization of an experience.’⁴⁰¹ Again, the case Hacker makes is a strong one. If we try to say that it is like V we are implicitly comparing V to something else. Something that we ourselves have done? Something that we have felt? Our *specific and subjective* qualia with another’s? What exactly are we trying to communicate by such? If we also say that it is ‘like’ something for A to do V then we are even more clearly off track as when A answers (‘It was wonderful/terrible’) the V in question is no longer *like* anything at all, it simply *is* (wonderful/terrible).

These objections, powerful as they are, cease to carry their conceptual weight when we understand qualia as types (although they may in certain instances retain some of their linguistic aspects): Hacker’s arguments seem to be against the framework of the ‘whiff of leonine’ version of qualia (the more common version), which I have also argued against. If qualia are instead understood as types then case (1) statements simply indicate the type referred to, and as types are at least species common (that is, typically similarly experienced by members of the same species (even though there might be slight variations within, as in the case of perception between a seeing and a non-seeing person)), a comparison between seems not only entirely apposite but indeed welcome and instructive. There is much that we might learn from each other if you describe and reflect on your experience of the pain of being pricked by a pin and I do the same for mine, and within

⁴⁰¹ *ibid.*, p. 166.

those descriptions there could be all sorts of ‘it is like’ statements that contribute variously and to varying degrees. Case (2) would also refer to the type as it is experienced by us, by all of us, and when A gives his ‘wonderful’ we understand what he means by referencing our own experiences of that type and what we made of them then (particularly when we too determined such to be wonderful). This comparison is once more perfectly valid when operating at the higher level of type. If A is saying that his thought was wonderful we know what it is for a creature such as ourselves to have a wonderful thought: we know the feel of it. Here we can maintain the determination of a connection involved with the emotional depiction because in speaking about experiential types we are not so much mixing these language acts (‘judgment of similarity’ and ‘affective attitudinal characterization’) as we are using them in mutual support. A ‘whiff of leonine’ reading of qualia would not permit this since the similarity bridge could not be crossed, indeed by definition – the ‘whiff of leonine’ definition – it would not even exist.

The other manner in which Nagel’s piece has been criticized is to attack its purported argument against physicalism; here I think that Nagel has simply been misread. This though is less of an argument against qualia and more of an argument regarding what might be underpinning consciousness (that is, awareness on our terms), and so with our lens drawn slightly out from the phenomenological level we have been at, let us continue. Yujin Nagasawa reads Nagel’s ‘bat’ piece as claiming that physicalism, if it be true, must be able to ‘provide complete explanation[s] of not only physical, chemical and biological but also phenomenal features of the world’,⁴⁰² but that as the bat argument shows it cannot it is therefore not true. Nagasawa puts it this way:

⁴⁰² Yujin Nagasawa, ‘Thomas vs. Thomas: A New Approach to Nagel’s Bat Argument’, *Inquiry: An Interdisciplinary Journal of Philosophy*, 46:3 (2003), 377-394 (p. 381). Nagasawa’s argument largely repeats an earlier one made by Frank Jackson in his ‘Epiphenomenal Qualia’, *The Philosophical Quarterly*, 32:127 (1982), 127-136.

(10) If physicalism is true then x , who knows everything physical about bats, knows everything about bats.

An addition of the following innocuous statement enables Nagel to derive the falsity of physicalism:

(11) If x knows everything about bats then x knows what it is like to be a bat.⁴⁰³

Since Nagel's argument arrives at the conclusion that we cannot know what it is like to be a bat without being a bat type creature ourselves, Nagasawa states, even being physically omniscient does not yield the necessary knowledge and hence physicalism must be false.⁴⁰⁴

I find this analysis wide of the mark. Nagel's argument against physicalism is that it is *incomplete*, not that it is false tout court. In his 'bat' piece Nagel writes, 'For there is no reason to suppose that a reduction which seems plausible when no attempt is made to account for consciousness can be extended to include consciousness. Without some idea, therefore, of what the subjective character of experience is, we cannot know what is required of physicalist theory.'⁴⁰⁵ This is clearly a thought directed against merely and straightforwardly extending an explanation which works on one level (absent accounting for consciousness (awareness)) to another level (including accounting for consciousness (awareness)), without adjusting that explanation in any way whatsoever. Comprehending how a bat's sonar works by studying the pitch of her cries and the biological apparatuses in her ears does not also tell us 'what it is like' subjectively to hear in the way a bat does. For physicalism to be complete, Nagel argues, it would need to also tell us that further datum. Yet it seems that to account for that physicalism would also need to be able to account for

⁴⁰³ Nagasawa *ibid.*, p. 381.

⁴⁰⁴ *ibid.*

⁴⁰⁵ Nagel 1979, 'What is it like to be a bat?', *op. cit.*, p. 167.

consciousness (at the level of awareness, this is referencing the experiential), which, aside from some assertions made in certain reductionist quarters, it does not appear at present – in its present form – to be able to do, at least not descriptively (although, as above, physiologically speaking the neuronal account seems sufficient). Does that mean that we need to reject physicalism? Not necessarily, for it does explain much and may only need to be fine-tuned a bit; a careful reading of Nagel indicates this is his position. That Nagel’s objections to physicalism in his ‘bat’ piece center on its current incompleteness, and not its simple falsity, is further reinforced by statements he has made to that very effect in other work done contemporaneously with the ‘bat’ piece and more recently.⁴⁰⁶ To read Nagel as making a case against physicalism wholesale is in my view to both misread him and to also risk failing to recognize the importance of the concept of qualia in the quest to understand consciousness in its fullness.⁴⁰⁷

In the previous section on modern neuroscience the physical basis for qualia was argued to be demonstrable through a consideration of the intuitive psychological judgments that our System 1 mode operates in and that are major constitutive components of our selves (through their reach into the three emotional sets): that is, the incessant, preconscious, and automatic analyses that the brain conducts on the input it receives as we go about our ordinary lives. I wrote there that, ‘There is thus no difference between what a feeling *is* and what a feeling *feels like*; there is only a difference in how we think about and discuss them.’ If we agree with Nagel that the current physicalist picture is incomplete, as I am inclined to, then how can we square the reductionist claim that ‘there is no difference’ with the

⁴⁰⁶ On the contemporary piece (first published in the 1979 collection referenced here (the ‘bat’ piece originally appeared in 1974)) see Thomas Nagel, ‘Panpsychism’, in *Mortal Questions* (New York: Cambridge University Press, 1979), 181-195; see also Nagel 2012, *op. cit.*

⁴⁰⁷ At the risk of overkill but with the intention of erring on the side of caution, please note that in writing about ‘what it is like’ the thinkers being considered are largely using ‘consciousness’ in the way we have used ‘awareness’; on our model wherein consciousness is the whole networked modular construct layers are employed that are not considered by others. The terminology can be confusing, and it is hoped that our shift to ‘awareness’ has provided a helpful parsing where none occurred before.

incompleteness claim that knowing something like the way in which biological apparatuses work does not grant the further knowledge of ‘what it is like’?

There are at least two ways to handle this, I think. One is simply to deny, as Hacker would (and Dennett, with others) that qualia are anything at all and determine the neurological picture I outlined to be sufficient, which is to say to remain only on the physico-chemical level of discussion. We found our gap to be closable on that level since we were able to locate a physico-chemical explanation for qualia; what then is the problem? It would seem that the physical picture is complete and that Nagel is simply wrong. That may be, but I am not entirely convinced, for what is missing from the physico-chemical explanation is the feel for us *of that explanation* – the feel that, even conceptually ‘knowing’ the why and wherefore of it, never quite disappears – and if we think about that we find another way to approach the incompleteness claim, and it may indicate not only an incompleteness but an *incompleteness*. The real problem is that we do not live, and we certainly do not mentally function, on the physico-chemical level; rather we move, think, breathe, on the conceptual level. That is how we are built, and since we are built that way a purely physico-chemical explanation seems highly unlikely to ever satisfy us. Does that mean that the physical picture is *incomplete* though? Does it not instead just mean that it is *unsatisfactory*? That hair could be split, but as our experience with revolutionary new ideas seems to show (moving from an earth centered view of the solar system to a sun centered one, for instance, or learning to accept Einstein’s relativity arguments, for another),⁴⁰⁸ we need to be able to intuit the view for it to really stick, and to do that we need to *feel* it. Perhaps in time the physico-chemical explanation of qualia given above will become as second-hand as the earth revolving around the sun has, but until then my own feel for it, at any rate, will

⁴⁰⁸ For an interesting discussion – one that can seem both dated and more applicable than ever – on this and related issues, see Feyerabend, *op. cit.*

remain that as good as it is it is not quite there yet. I suspect that I am not alone in this judgment (excluding Nagel's present company, that is).

One other point on this issue: William G. Lycan has suggested that introspection works on stacked second-order representations of first-order psychological states, and that the tokens we use in introspection have both referents and modes of presentation; qualia, therefore, do indeed provide phenomenal information that is unavailable to 'third-person science' even though no 'special phenomenal facts' are to be found.⁴⁰⁹ This, we can see, is another way of describing how the feel of an explanation differs from the explanation itself, and how that feel, for better or for worse, does seem to be an inescapable part of us. It is interesting to note too that Lycan's description of ordered representations appears to match with the neuronal map-making picture given previously.⁴¹⁰ It could well be that the physico-chemical is all there is – we just do not live that way.

I have argued that qualia should be considered as types, that the case made against physicalism is not that it is false but that it is currently incomplete. I have also suggested that we ought to regard ourselves as always having consciousness – though differently active and to varying degrees – as long as we are alive, and that consciousness is an important compositional element of the self. The self, moreover, and especially on our definition of it, resides on that very conceptual level where we find physicalism's incompleteness to be – the level, again, where we spend all or nearly all of our thinking lives. The implications for a realist self of such reflections are far-reaching and will be covered. Nearer to hand though, and in order to pave the way for those further explorations of a realist self, the manner in which the physico-chemical generates qualia – whether or

⁴⁰⁹ William G. Lycan, *Consciousness and Experience* (Cambridge, MA: The MIT Press, 1996), pp. 100-101.

⁴¹⁰ In the above; on map-making see especially Damásio 2012, *op. cit.*

not we agree on the sufficiency of the entirely reductionist view – raises a new question which is often asked at present: Might it be that we can build a computer which works enough like a brain to be generative of its own emergent consciousness and hence also self? The simple answer to that question, I think, is ‘No.’ The fuller answer makes up the next portion of our investigation.

B. Artificial intelligence and the self: The importance of qualia

To begin to discuss this issue a clarification might be in order regarding the subject itself, notably a reminder that consciousness does not equal intelligence and that it is quite possible to have intelligence – artificial or natural – without consciousness: as an example we need think no further than a common calculator. Yet amidst the media hype surrounding the coming explosion of artificial intelligence those who are speaking about the topic are generally not referring to this limited sort of intelligence, rather they mean to discuss *intelligence that is self-aware* and that, on the basis of where our analysis has brought us and as far as we can currently understand, must include some form of consciousness. By starting to now also bring in the further conceptual baggage of ‘intelligence’ we run the risk of blurring dividing lines that need to be kept clear, and so to explore this topic we will initially set out some thoughts on intelligence proper before making our way back again to consciousness, and from there to its application to the self.

First of all is the issue of brain complexity and measured/measurable levels of intelligence. Until very recently researchers in the field thought – were convinced – that a large and complex brain including a neocortex or a neocortex-like structure (a part of the higher brain that regulates sensory perception and language) was necessary to distinguish between human faces. As all humans have the same basic features (two eyes above a nose and mouth that are more or less centrally featured within an oblong shaped head), it had been

considered that only primates were up to the complexity of the task.⁴¹¹ It has however recently been established that birds, with their neocortex-like structures (but not proper neocortexes) can do this too, yet surprisingly also that the archerfish – a tropical species that like all fish has only a simple brain wholly lacking in anything like a neocortex – can do as well.⁴¹² If even a fish is capable of what had heretofore been considered the domain of the most advanced mammalian group on the planet, then either we do not really understand intelligence or we do not really understand how the brain works; or perhaps rather both. If that is so, if we do not in fact have the grasp on the biological brain that we thought we had, then what hope might there be of building an artificial intelligence that could successfully mimic or even outdo the brain?

We have great hope, it might be objected, for not only have we been doing so already but the archerfish example shows how little intricacy is really needed for higher abilities to emerge. After all, computers running artificial intelligence programs have sometimes beaten humans at chess, and even famously at the much more demanding *go*, the world's oldest and most strategically complex board game. We are teaching our machines to be able to do all sorts of things that were once considered unimaginable, it is said; surely the sky must be the limit as our computing technology advances further and further.

I think though that we must give pause to such speculating because there are all sorts of theoretical confusions threatening to creep in, and again that largely as a result of the language used in discussing these matters, stemming as it does from the tendency to conflate the physico-chemical with the conceptual. A large step back is required, and with

⁴¹¹ AFP-JIJI, ““Smart” tropical fish can recognize human faces’, *The Japan Times: On Sunday*, 26 June 2016, 21.

⁴¹² This type of fish naturally sprays water at insects and so the individuals used in the study were trained to instead spray water at a photograph of a specific human face and were then shown a number of different facial photographs to test whether or not recognition was achieved via only spraying water at the ‘correct’ face’s photograph; *ibid.*

it a reassessment of point of view. To start, and contrary to the just considered assertion, no one is ‘teaching’ a machine anything, programs are simply being written which will then be automatically run. When computers execute these programs nothing at all is occurring on any kind of metalevel, on any kind of conceptual level. We might think so, but if we do that is because we have a tremendous tendency to anthropomorphize the things we build, especially when those things operate more or less on their own. In another instance of this habit it has recently been reported that robots are being instructed to ‘feel pain’.⁴¹³ In such cases the robot does not ‘feel pain’ in anything like the way that you or I or even an archerfish feels pain: what the robot *feels* is an absolute blank, the device is simply alerted via installed sensors to the fact that it is in a situation which might damage it and so it ought to move. Yet, it may be asked, is that not exactly what is happening to us when we feel pain? Not at all, for what is missing in the robot’s case are the qualia associated with pain, the way that pain is not only processed but experienced, as we have been expounding. That higher abstract and conceptual way involves – or inescapably seems that it involves – more than the physico-chemical foundation of it, and again that seeming is why I am far more willing to agree with Nagel’s charge of the physicalist picture being incomplete than I am to disagree with it, even if I accept that there is an explainable physical basis for qualia in the manner described.

Artificial intelligence programs (whether within or without robots), for all of their impressiveness in problem solving and even creative problem solving, have nothing happening on any meta, any conceptual, level. They experience no qualia, and therefore if we can say anything about their awareness at all what seems clearest is to say that they have no awareness and stop there. In reflecting on this it does now seem appropriate (on

⁴¹³ ‘Researchers teach robots to “feel pain”’, *BBC News: Technology*, 26 May 2016. <<http://www.bbc.com/news/technology-36387563>>.

our understanding) to also equate the having of qualia – interpreted as type – with consciousness, but only at its layer of awareness and not the entirety of how we have considered consciousness in our study; this ranking of qualia with consciousness as awareness was the same definitional move we saw Nagel and others making above (remembering that on their usage ‘consciousness’ is employed primarily along conventional lines to indicate ‘awareness’ and nothing further). The connection moreover seems particularly appropriate if we also take consciousness as coming in degrees, as was argued, and thereby grant that associated conceptual modes too differ by degree (recalling again the discussion of more and less complicated neural maps). In pulling all of these strands together we appear to be making some welcome progress.

However the temptation remains to say that if consciousness – of whatever status – can emerge from a structure as simple as some of the brains we see in living organisms, far simpler even than a fish’s, then surely a sufficiently wired ‘electronic brain’ could give rise to consciousness too. (Panpsychists in particular would have trouble denying this as their position holds that consciousness is universal.) Perhaps it could at the level of the underlying network: employing artificial preconscious System 1 decision-making or the like? Yet would a label such as ‘consciousness’ apply without a comparative awareness? What does it imply that nonhuman animals who have brains only equipped with System 1 analytical tools (e.g. our archerfish) still have awareness? The case is far from being clear cut, and for that fact should probably be left open, but at present we can say that we have yet to see anything like consciousness happen in a machine – again, even allowing that there are different degrees of it –, and until it does perhaps the best that can be claimed is that what is really not understood is emergence itself. Whatever process is involved in the arising of full consciousness out of a functioning organic brain could conceivably be beyond our ability to grasp and/or imagine, but it does at least seem that qualia, and

therefore the experiential functioning of an abstract conceptual level which is in whatever way 'beyond' the purely physico-chemical, needs to be present.

For the self to exist consciousness must necessarily be a basic element of its make-up, a point repeatedly argued for here. If artificial intelligence programs are never able to demonstrate consciousness (fully, in all its layers), therefore, such will similarly not be able to have the self-awareness that some technologists have been dreaming of. In my view the goal of building an artificial intelligence that is self-aware is probably a fundamentally flawed one, and it is so because of qualia, because there is never anything that it is like to be a machine. To 'be' a computer is simply to act out internal programming, to remain on that unaware and non-experiential physico-chemical level. This point is reinforced if we entertain the notion that even if we were to adopt a fully determinist view of the universe in which we and all life functioned in the same way of blindly following internal programming there would still be that difference: although our computers and us were powerless and without any free will (this being the hardest determinism available), we would yet be feeling (qualia) our way through our preprogrammed steps while the computer would not. Without that feeling all bets for something more, something emergent, appear to be firmly off.

C. Software, hardware, and wondering about the real

Finally, let us now consider an analogy about the self that, in the midst of the current climate celebrating the growth of artificial intelligence and with the so-called Digital Revolution firmly established in our cultures and our minds, seems to provide much clarity. It appears to be so fresh and new – despite having somewhere around three decades of intellectual history behind it – partly because it was not made well known until Ray Kurzweil published his *The Age of Spiritual Machines* just prior to the turn of the

millennium,⁴¹⁴ and this supposed novelty adds to its luster. Given these contributing factors, and particularly with the contemporary prevalence of anti-realist accounts in academic and other circles, the comparison has even become something of a commonplace, mainstream almost without question. The analogy, of course, is that between the mind and the brain on one hand and software and hardware on the other; or, in its more self-oriented version, between the self as software and the body as hardware. (Already here we can see the old dichotomy of self equated purely with mind.) A human being, we are told, is like a personal computer: we have our hardware bodies and we have our software selves, and it is thought that this analogy works because software is virtual: it functions without really existing, just as the self is ever with us without really being there at all. Could this possibly be right? What does it mean to say that software (and the self) is virtual?

A claim along these lines is first to assert that nothing physical exists which could be pointed to as the object under consideration, we are instead only able to indicate evidence of the object and thereby infer its presence (of sorts). On the face of it this looks rather nonsensical; software can after all be bought from the store and taken home and inserted into a computer to be installed. Yet software can also be downloaded and installed that way: in that case what is the physical object? Is there one? If not, where is the software? When people say that software is virtual they do not mean that it does not exist in the world; they mean rather that it is not something one can reach out one's hand and hold onto – note that the plastic disk used to put the latest version of Microsoft Blah into a device is not in fact the software itself, it is merely a delivery vehicle for the software. What then is the actual software? It is code that allows a computer to run in a way that it theretofore

⁴¹⁴ Ray Kurzweil, *The Age of Spiritual Machines: When Computers Exceed Human Intelligence* (New York: Viking Press, 1999); for an earlier example of the representative thoughts behind the analogy see also Scott Harrison, 'On the Analogy Between Mind/Brain and Software/Hardware', *Massline.org*, 04 December 1992; last update 24 June 1998. <<http://www.massline.org/Philosophy/ScottH/mindsoft.htm>>.

could not, at its most basic level it is a string of ones and zeroes. We again confuse ourselves if we think only on the conceptual level of human-readable code that is the higher rung at which software is written by and for the maintenance and engineering of programmers (recognizable words such as ‘run’, ‘frame’, ‘printf’, ‘main’, et cetera). This level of code is translated into machine code to obtain the instructions for the central processing unit, the letters and numbers of which ultimately stem from binary sources (bit strings) which represent them. All of this is of course built into the computer – that is, built into the computer’s *hardware*. The software we have been wondering about is therefore merely the coding that shifts how the hardware functions, and although one can pick up and move about the various pieces of hardware that are put together into the physical objects on one’s desks and in one’s pockets, one cannot so pick up the software (although again one can pick up the medium by which the software is transferred from place to place). It is in this way that software is virtual, that it is considered to be physically nonexistent.

We must stop at those last two words, for it is here where a deep confusion threatens to set in, confusion about the self, about consciousness, and about a great many things. We mistakenly tend to think that the nonphysical is *not there at all*, that an absence of physical presence means an absence period: empty nothingness. This assumption may be the result of an (reflexively? acculturated? historically-bound?) uncritical examination, or it may just be how we happen to be tuned, biologically speaking. If there is nothing material then we say there is nothing at all; yet we do not say software is nothing at all, nor do we say that software is a mere illusion that stems from the natural way the hardware functions (as some of us do about the self).

To fill in this space between, so to speak, it has been suggested that software must first be encoded (i.e. written down) in order for it to exist, and those who make such claims hope

thereby to provide at least some degree of materiality.⁴¹⁵ Yet would we say that the software exists while its commands are still being thought about by its writer as she types? Does she need to actually hit the keys before the software shifts from being non-existent to existent? Lines could be drawn all around these questions and things have now become very interesting. Let us continue.

Software depends on hardware for its realization, and software functions – its existential purpose – in order to alter how the hardware that is its foundation and target operates. Might this be an allusion to dualism? It is and it is not; software is not something that could function separately on its own, although we might be tempted to say that it could exist separately on its own, for after all uninstalled software surely seems to exist in some manner. Lycan makes a similar point when he argues against making any distinction between ‘software’ (speaking broadly, the context is in reference to the brain) and ‘the hardware it runs on’, which he further puts as a – equally erroneous and to be avoided – separation between ‘the functional’ and ‘the merely physiological’. He writes that, ‘The difference between physiological functional talk and more abstract functional and computational talk is just that, a difference of degree of abstraction and functional organization.’⁴¹⁶ This echoes our above comments on consciousness and the levels on which we discuss it (and hence also think about it), and how it might work. Granted, no software would exist anywhere without the requisite hardware already being in place, but this is not a chicken-or-the-egg query and that is not the point. Software that has been written but has not yet found a home, as it were, is something that appears to exist in the way that we might think a disembodied soul exists: simultaneously there and not there, doing something but doing nothing, *being* something while being nothing. Is that what

⁴¹⁵ See, for example, Harrison, *ibid.*

⁴¹⁶ Lycan, *op.cit.*, p. 118.

those who compare the body to hardware and the self to software are really trying to say? Pulled out in this way we find ourselves arriving not at the non-realist destination we supposedly set sail for when we embarked on this analogy but rather at a very Cartesian-looking port. By trying to argue that the virtual is real without being real, causal without being material, proponents of this view (and here I would include non-realists about the self) have painted themselves into a corner. They must either admit that the immaterial can exist in a very real way, or they must assign some version of materiality to software (such as insisting that it first be physically written), at which point the non-realist self comparison that was sought after fails.

I do not think the thought behind this analogy has gotten that far, however. It has instead stopped at the point where it is taken that a suitable means of understanding how the self is actually nothing has been achieved. In my estimation, the argument tucked away in this analogy is meant to go no further than to state that although one may think the self really exists based on one's personal experiences of it in fact it does not, and moreover does not need to. It is not my view that those who make this claim wish to further argue that software does not exist (some might, but I doubt it); although it does seem that by this analogy/argument we must draw that conclusion anyway for if software can exist virtually then the self can exist virtually too. This is a tantalizing thought for a realist claim about the self, but let us briefly table it to linger a moment longer on the issue of software existing only 'virtually' because I think that there is still a misconception present that we have not yet rooted out, namely, software does not *exist* virtually it *functions* virtually.

Software itself is code, and readable to us or not code is a form of text and it would take quite the hardy solipsist to argue that text does not exist (but, being a solipsist, that the self nevertheless does?). This code need not even be printed out or written down to be said to

exist, for surely most of us would agree that language exists in some sense even if it is only – and only ever has been – spoken. This judgment sheds light on our earlier query regarding at what point, as our programmer sits over her keyboard, the code she is thinking to type comes into existence *as* code. Software, it seems, must be said to really be there amongst the hardware. Yet how? In what way? I think in dwelling on this subject we are approaching something quite remarkable. What these thoughts appear to be leading towards is the conclusion that there is a middle road between the physical-material and the ideal-immaterial, and software gives us a clue as to what that might be. This middle road is the *functionally existent*, which we might note is a type of emergence. It manifests itself in its operation but ontologically we have trouble placing it within our standard categories. Lycan again offers a refreshing insight on just this issue. He introduces what he terms ‘intentional inexistents’, objects or representatum that are internally given as ‘physical’ objects but are really more in line with concepts. The important difference here, if I read him correctly, appearing to be in how we perceptively understand and think about (talk about) versus an actual empirical (physico-chemical) state. Our levels again. Lycan gives the example of ‘the sky’ as a thing, a blue thing, to flesh out what he means:

no one thinks that there is any nonexistent *physical thing* up there. But I contend that there is a nonexistent physical thing up there: It is an illusion. Vision represents ‘the sky’ as an object. Poets write about it. Visually, the sky is *a canopy*, or the *vault* of heaven.⁴¹⁷

This is highly instructive, and it demonstrates just how much weight an idea can have for the kind of abstract and conceptual creatures that we are: an idea can be real in many more ways than purely empirical considerations would limit us to. This is precisely the kind of approach that will be explored in the first half of the next and final chapter, where we will

⁴¹⁷ *ibid.*, pp. 152-153, quote taken from note 9 on p. 189; emphases in the original.

focus on the possible metaphysics of a realist self.

5.3 A query at arrival

In our considerations on software and hardware, following on as they did from the examination of artificial intelligence and the question of a machine self, we can draw on the notion of text to ask one more enlightening question: Does a computer read software in the way that we read a language? That is, even at the root binary level, does a computer approach a string of zeroes and ones the way we approach a string of vowels and consonants? Based on our conclusions regarding artificial intelligence above we will likely answer this in the negative, and I think we are right to do so, for when a computer is ‘reading’ code there is absolutely nothing happening at the metalevel, at the qualia level, whereas for us when we are engaged in reading all sorts of things are taking place at the metalevel, and necessarily so. Additionally, pushing on these analogies even further, we can say that all of our metalevel happenings – all of our qualia, all of our concepts – too function virtually, and in a way that is existentially deeper than the way in which software functions virtually. We are not only unthinkingly zipping through instructions, we are engaging, considering, weighing, reacting, emoting, analyzing, each and every time and indeed all of the time. A comparison such as the one we have just been analyzing is not only inaccurate it is wrong, misleading about both of its component parts: software and the self. Whatever the self may be it is a part of us in a way beyond that in which software is a part of hardware, and if software can be said to exist on its own and in a real (if immaterial) way, then it stands to reason that similar – and even more profound – claims could be made for the self as well. We are now in a position to consider just that, and our study’s closing chapter will pick up there.

Chapter 6: Metaphysics and Time: The Reality of a realist self and its (re-)making

6.1 Metaphysics for a realist self

In this opening section of our final chapter we will aim to establish in what sense the self concept that has been argued for could be considered to be ‘real’. We have already rejected such accounts as the Cartesian and traditional soul views, which were termed ‘hard realist’ selves, and opted instead for a ‘soft realist’ approach, yet as the reader will have noticed we still – after five full chapters – have not provided any remotely satisfactory definition or exposition of what such a soft realist self might actually be in regards to its realism. Indeed, our commonsense understanding of reality likely balks at the very idea of soft realism. Still, the present study’s failings notwithstanding, we have made good headway towards establishing an alternative view of the self that sees it as something which is very much a permanent part of us, specifically and uniquely so, but that is also constantly undergoing change as our paths through life unwind, a self that is established from both external and internal forces, but whose ‘making’ – whose creation – is likewise never-ending. The picture of what is being proposed here has hopefully become clearer as we have progressed, and so let us now face the issue head on and confront this ‘real’ self, taking as our starting point the concept of functional existence as it was examined in the previous chapter.

To think about the possibilities that an idea like functional existence opens up we need firstly to question the unexamined position on what it means to exist and what it means to be real, for the conceptual backgrounds of both existence and reality have been heavily influenced in contemporary times by outlooks imported from the assumptions of the hard sciences. We have been taught and grown accustomed to presuming that only what is

empirically, materially provable is valid, and as such have severely reduced the ideational space in which we allow ourselves to operate. Unfortunately, the exit from this mental cul-de-sac presented by modernity's approach to empiricism is difficult to find,⁴¹⁸ but some steps can be made in that direction, and they will prove sufficient for the self theory that we have been exploring.

To start to bend, to loosen and free, our thinking along these lines of what existence and reality entail, consider the following scenario that Peter van Inwagen asks us to entertain: A group of soldiers using bulldozers in a desert builds a fort made out of sand; have they brought anything into existence?⁴¹⁹ We are inclined to answer that of course they have, they have brought a fort into existence, and the evidence for it is as plain as day and standing there in the structure now visible, touchable, even measurable, before our eyes. van Inwagen argues against this conclusion, stating instead that really they only 'rearranged the furniture of the earth without adding to it', although he does admit that 'if one of the legionnaires said, "We have built a fort," he would thereby assert a true proposition.'⁴²⁰ According to van Inwagen, to presume that there is a real fort in existence is to contend that there are two distinct objects which are fully spatially coincident (the sand and the fort) and that differ only in modal and historical properties; he finds this incredible and labels the fort a 'virtual object'. He claims that we only speak of the persistence of artifacts in the way we do because of the covert references to intelligent beings' activities, to histories of maintenance, that are latent in the words we use.⁴²¹ Are we really so deceived by our own languages?

⁴¹⁸ Paul Feyerabend has provided an excellent guide with his classic *Against Method*, op. cit.

⁴¹⁹ Peter van Inwagen, *Material Beings* (Ithaca, NY: Cornell University Press, 1990).

⁴²⁰ *ibid.*, p. 124.

⁴²¹ *ibid.*, see especially the discussions on pp. 125-127 and 134-135.

We may not find this line of reasoning wholly satisfying, nor may we find it altogether clear in either an empirical or an unempirical sense. On the one hand the soldiers, if they have only used sand in building their fort and not added any other elements such as supporting beams to frame the walls, indeed did little more than reshape present materials into a new (but previously potential) form, while on the other hand we are surely correct in our use of descriptive language (as van Inwagen himself admits) to say that the soldiers have ‘built a fort’. Where does the empirical evidence here lie? On which side of the fort/no fort question (or ‘object/virtual object’ question, to put it in van Inwagen’s terms) does it fall? Since the fort is visible, touchable, measurable – all aspects of empiricism’s standard operating procedures – does that make its existence scientifically conclusive? Or does the fact that the fort is composed of no other visible, touchable, measurable constituents than the same sand as that which surrounds it make its nonexistence scientifically conclusive? How would a good empiricist approaching this edifice respond? We might think that he would agree with our initial assessment, that the fort is ‘really real’, that it exists, that naturally it exists, and that is why we so easily and assuredly speak of its existence. van Inwagen’s point is a subtle one though, and it does make us reconsider our position when we admit that compositionally there is nothing anywhere but the very same sand (ignoring subatomic particles and whatever structural variation there might be amongst the grains of sand at that level). Is the issue not then one of our conceptual perspective? We need not agree with van Inwagen that the fort is a ‘virtual object’, and we might wish to argue that the fort exists far more straightforwardly than our earlier example of software does (or we might agree and might not wish to so argue, either way), but whatever our reaction to the query van Inwagen gives us, in facing it we find ourselves sufficiently shaken to now be able to move into a consideration of the self that places the real in a less well-defined position than we are used to. Our thinking has opened up.

Graham Harman takes us further down these untrodden paths via an examination of Heidegger's famous tool analysis in his *Being and Time*.⁴²² Harman first situates us by expanding our understanding of the term 'tool', which given its usage in everyday terms has strong nuances (and therefore buried assumptions) of 'physical thing'; he reminds us that 'Equipment is not effective "because people can use it"; on the contrary, it can only be used because it is *capable of an effect*, of inflicting some kind of blow on reality. In short, the tool isn't "used" – it *is*.'⁴²³ As such, equipment/tools therefore ought to be understood in the same way that any entity that can exert itself or its force upon an inhabited environment is understood, and we can conclude, I think, that surely concepts would fall into this category: ideas inform views and outlooks, views and outlooks inform attitudes and beliefs, attitudes and beliefs inform actions. The power of the conceptual lies not just in description, nor even in its linguistic consequences, but in the 'world' it builds (i.e. in the sense of Husserl's 'lifeworld' and/or Heidegger's extension of that) through the comprehensive structuring of its holder(s). If I take it that X and truly believe X to be accurate then it will become a part of the embedded nature of my existence out of which I move, and the behavioral results of that stance will flow quite naturally from there. Moreover, given that in our empirically-driven times 'accurate' is equated with 'ultimate truth', the consequences of this perspectivism will be profound indeed, some of which we can see every time we open up a newspaper and read how our leaders relate and form their policies, to say nothing of our everyday interactions on the street with those from differing cultural or ideological backgrounds. If we apply these thoughts to the self we can appreciate just how profound and how potentially revolutionary they can be.

⁴²² Graham Harman, *Tool-Being: Heidegger and the Metaphysics of Objects* (Peru, IL: Open Court, 2002); Martin Heidegger, *Being and Time*, trans. by Joan Stambaugh, rev. and for. by Dennis J. Schmidt (Albany, NY: State University of New York Press, 2010).

⁴²³ Harman, *ibid.*, p. 20; emphases in the original.

The self-view we hold, whether reflected upon or not, is an idea that has far reaching influences on every aspect of our lives through its input on our derivative views, beliefs, and actions; this judgment is nothing new, and is moreover obvious enough, I think, to allow us to move directly to the next inquiry, which is this: The self is a concept whose potency cannot be denied, fair enough, but does that make it *real*? Is not the point rather that whatever self concept is adhered to will be generative or partially generative of certain conduct, and that the measurable results of those actions when performed in surrounding environs will be what is/are real? *What makes something real*? This is our core question, and it harkens directly back to our findings on software and the functionally existent.

Harman writes:

Anything, prior to erupting in its explicit form, is real simply by exerting its efforts in the cosmos, by breathing its life into a world that would not have been the same without it. In other words, before any object is present-at-hand [*Vorhandenheit*], it is ready-to-hand [*Zuhandenheit*]: sincerely engaged in executing itself, inaugurating a reality in which its characteristic style is unleashed.⁴²⁴

Dare we begin to speak of the real as the (merely) influential? Of the substantive as that without any necessarily physical or measurable substance? Of the unempirical as the actually present? If we do then we have embraced the functionally existent, for that is precisely what is on the table here. We go deeper.

The functionally existent self is a concept that can be hunted for, and although it has perhaps often gone unnoticed once identified it demonstrates firm and stable historical roots. It is characterized by the position that a person's self-view is not only expressive and

⁴²⁴ *ibid.*, pp. 219-220.

exhibiting, but is far more than what is externally manifest and observable: it is even structural, constitutive, compositional, and in that individual building. It is the idea of the constructive, creative power of an idea, and once found it must be confronted. In a study on G.W.F. Hegel's Chapter Four in *The Phenomenology of Spirit* ('The Truth of Self-Certainty'),⁴²⁵ which is focused on self-consciousness and unequivocally places what we have called the 'functionally existent self' front and center (though without of course using those terms), Robert B. Pippin writes of the Hegelean position that:

how I take myself to be is *self-constituting*; I *am* who I take myself to be or can only be said to be an I or subject insofar as I determinately take myself to be such and such, in some determinate way or other, and I accordingly functionally vary as such self-constituted takings vary.⁴²⁶

This explication alone is quite interesting and appears to fit in well with our concerns about the self, although without covering the whole of them, nor our full self-view. There is not any necessarily persisting 'I/ego' in the Cartesian sense to be found here, but there is nevertheless an 'I/ego' that is real and substantive enough to be constitutional of a self in a way that is directive and world engaging. This is a self that is conceived and composed and then thereafter sustained, maintained, or tweaked, adjusted, shifted, as circumstances differ, and it is furthermore one that is carried over time. In addition to Hegel's stance we may offer that given the material covered above in our study such alternations would almost certainly happen at preconscious levels as the individual moves and acts within their environment, dealing with the multifarious pressures and concerns with which we are all

⁴²⁵ Georg Wilhelm Friedrich Hegel, *The Phenomenology of Spirit*, trans. by A.V. Miller, for. and comm. by J.N. Findlay, rev. edn (Oxford: Oxford University Press, 1977).

⁴²⁶ Robert B. Pippin, *Hegel on Self-Consciousness: Desire and Death in the Phenomenology of Spirit* (Princeton, NJ: Princeton University Press, 2011), p. 68; emphases in the original.

daily familiar. Hegel's work as elucidated by Pippin is very concerned with the other in all of this, and if our deeds as seen by others do not match the projections we give them it may be, Pippin writes, that I 'turn out not to be whom I took myself to be' – that is to say, the self ascriptions that I have internalized would not be the same descriptors another would use when defining what I am like – but nevertheless 'that erroneous self-conception [i.e. the one I hold despite not being (from another's point of view) 'whom I took myself to be'] is still an essential dimension of who I am.'⁴²⁷ This outcome ('still an essential dimension') is important and noteworthy. Thus, even in cases where there is a mismatch between how I take myself to be and how the people around me take me to be, what appears by this to be most pertinent for my own self, my own built and building, made and making and functioning self, is that view from within, that notion of self which I hold and no doubt more often than not cling to dearly. On our model many of the relevant elements would be included in Set 1, and from that aspect act towards the full additive self as we have defined it.

The importance of self-reflection here is quite clear. In that we also find a caveat on available insightfulness though, for as Christopher Peacocke points out, bodily (proprioceptively) knowing that one's legs are crossed does not support a conceptual knowledge of what it is/what it means to have one's legs crossed, just as 'The ability to represent [mentally] "they are looking at me", and "they are doing such-and-such in relation to me", do not imply the ability to gain knowledge about oneself from a third person perspective on oneself.'⁴²⁸ This is naturally not to say that learning about oneself, *one's self* (in whatever realist or non-realist version), cannot be accomplished from outside sources, rather that the first-person perspective will always be foremost, and that what

⁴²⁷ *ibid.*, p. 68.

⁴²⁸ Christopher Peacocke, *The Mirror of the World: Subjects, Consciousness, and Self-Consciousness* (Oxford: Oxford University Press, 2014), p. 262.

learning does take place will come in differing degrees depending on individual factors such as thoughtfulness, acuity, motivation, et cetera. What role, we may ask, does the conceptual, does the epistemic, play in all of this? What *is* a functionally existent self, what kind of realism does it have, and what does this mean for self theories proper?

Given the overall academic context, the currently dominant anti-realist stance is probably a fair response as far as an answer to the hard realist traditional soul or Cartesian self conceptions are concerned – certainly it has done much to move the discourse beyond Cartesianism –, but at this point in humanity’s five thousand year-plus discussion on the topic of the self it is difficult to be sure about anything. However, what contemporary non-realist self-theories have missed, I think, is this (other) level of function, and the pivotal role that it plays in our lives as we experience them and when viewed from within. If our own psychological understanding of ourselves – of what we are – is partially constructive of an actual self (including such items as inherited traits and along with consciousness and bodily presence), as it has been argued to be in the foregoing, then this conceptually based, created and carried, formed and maintained, internal abstract posit has a degree of reality that cannot be denied: a working and participatory reality. Although Kristjánsson does not define it in any great detail, this is perhaps what he was getting at with his description of the self as an emotionally-based day-to-day psychological unit.⁴²⁹ This self is played out continually, expressive in each action and background to each thought and feeling. It is a functionally existent self, and it is foundational to a conscious and aware creature’s being – though we must exercise caution in a claim such as that as surely the complexity of the proposed self would differ widely with the degree of consciousness and hence too awareness (e.g. a dog’s ‘self’ in this sense would have far less

⁴²⁹ Kristjánsson, *op. cit.*; see also this study’s Chapter 2 Section 2.2.A (‘Psychological issues and the self: Cognitive structure’).

depth than a dolphin's, and similarly when compared with a human's; the reader is reminded that in the present our focus is solely on the human case). Yet what kind of substance might a functionally existent self be said to have? If it lacks the full robustness of a hard realist self then what could it really consist in? To answer that I think we need to begin by loosening our view on objectivity,⁴³⁰ for demands related to 'objective proof' that are based in unexamined empirical assumptions will present major obstacles to any progress we may make here.

To do so let us follow van Inwagen's lead and pose a query: What color is the sky? The question is beyond ridiculous; blue, of course! As the words leave our mouths we are forced to pause though, for is the sky always the same blue? Is its blueness identical for you and for me? What about for the color blind? What about when it is very cloudy? Is the sky's color ubiquitous, or does it depend on the ocular functioning of the viewing organism? Moreover, as Lycan made us think at the end of the previous chapter, what is the sky anyway? If we speak (think) at the physico-chemical level it is not even a 'thing', really, it is a collection of stratified gases, and in that it is not 'the' sky either as its compositional multiplicity means that that definite article has no business being connected with it. If we speak (think) at the conceptual level, however, then quite clearly it is a thing, a singular thing, that beauty up there above our heads upon which we gaze in wonder at times while at others we shake our fists in anger. Can we settle the tension between these two sides? We might strive for a more far-reaching description, perhaps by exploring its layers using the tools of physics and examining its subatomic particles – but then, as is by now infamously well known – any readings taken from the subatomic level affect and alter

⁴³⁰ For a fuller exploration of this idea see my short article 'Thinking Unempirically', *Philosopher*, 03 July 2017. <<http://philosopher.io/Thinking-Unempirically>>. Many antecedents to the ideas discussed in the below can be found in Heidegger's work: see the issues covered in Chapter 3, Section 3.3.D ('Phenomenologically-based accounts and related issues: Phenomenology 4: Heidegger on the self'), aptly illustrated, I think, by the distinction highlighted there between the notions of "'the" world' versus "'a" world'.

what is being explored, giving a necessarily distorted output that must be (imperfectly) mathematically adjusted for afterwards. Additionally to that modified data we will then add our own interpretations and introduce a further degree of variance and potential error. What has happened to us? Our question was of the simplest sort and yet we find ourselves utterly confused and incapable of even the most guarded and restrained attempt at a global answer.

Another example: I have a glasses case on my desk; I measure it and find it to be roughly ten centimeters long. Is that objectively true? Naturally – but only in a sense. What is important, I think, is that behind the concept ‘centimeters’ lies nothing but a standardized way of measuring that the world has inherited out of the brilliance of the French Revolution. As remarkable as the system is though, there is nothing ‘ultimate’ underlying it; there is, rather, something fairly arbitrary: the first meter was marked as one ten millionth of the distance from the North Pole to the Equator when measured as running through Paris.⁴³¹ Is that distance a fixed mark of the cosmos? Is it really unchanging in perpetuity throughout the universe? Meter measurement has of course been adjusted since then, but the questions of unceasing stability and permanence still seem to apply, and with them our wondering about any sort of ‘pure objectivity’.

What an analysis along these lines demonstrates is the unavoidable perspectivism and history that is carried by each concept and methodology we employ. What we determine ourselves to be seeing depends very much on how (in a broad sense) we see it, and what we conclude hinges entirely on the way we have been made to think; not least by the educations we have received, but also by sociohistorical settings, cultural backgrounds, even the languages that we use with their inherent networks of mental associations: the

⁴³¹ For a brief history of the conventions and the position they hold today, see ‘Metric system’, *Wikipedia: The Free Encyclopedia*. <https://en.wikipedia.org/wiki/Metric_system>.

embedded aspect of human life that this study has repeatedly stressed.⁴³² Yet our discussion here is not meant as an appeal to relativism; rather it is meant to highlight the unexamined position we often take in regards to objectivity and the insistence we often bring to all manner of endeavors that they must demonstrate and maintain a certain scientific rigor. The default stance concerning the theoretical today, be it metaphysical or otherwise, seems to be an automatic association of what is ‘objective’, ‘foundational’, ‘ultimate’, ‘true’, with the ostensibly empirically provable – and that only. What I wish to argue here is that as powerful and as clearly effective as quantitative research has been, it cannot, and should not, be asked to carry the entirety of human incidence on its shoulders, and if it cannot do that then it similarly cannot be held to account for every phenomena currently known and yet to be discovered. At least, it ought not to be so initially, and room should be left in our understanding and our ideascapes for that which is not scientifically explained but which nevertheless appears in our lives and our experiences and is taken to be relevant and significant to us. What I am suggesting is that we not only acknowledge but embrace the flexible and limited condition that we hold as the knowers we are, tucked away in our little corner of the Solar System, and in that open ourselves up to an appreciation of the world that starts from the experiential and takes that to be significant simply as it is: as felt, as lived. Can there ever really be an entirely objective, unbiased, and everywhere applicable response to anything? If we accept the above then we are led to conclude that such is indeed an impossibility, and if there is nothing without perspective then there can be no absolute objectivity, hence to build our lives and to base our thoughts

⁴³² See especially the discussion in Chapter 2 Section 2.5 (‘Certain uncertainty, randomness, and limited choice’). A further instance of this: traffic lights in Japanese are said to go from red to yellow to ‘blue’, despite the actual color involved following the international convention of being green. This linguistic quirk stems from the linkage of newness or inexperience or ‘greenness’ (as the notions are connected in English and many other languages) with the word for ‘blue’ in Japanese. See the following article for a more general discussion and some helpful examples: Peter Backhaus, ‘The Japanese traffic light blues: Stop on red, go on what?’, *The Japan Times*, 25 February 2013. Available on: <<http://www.japantimes.co.jp/life/2013/02/25/language/the-japanese-traffic-light-blues-stop-on-red-go-on-what/#.WUhlbISGOUk>>.

– to define and delineate, hem in and wither – exclusively on what we take to be objectively justified because ‘provable’ is to cut short our existence and to restrict our world, it is to stifle our very possibilities for being.

If there is no purely objective position from which we might argue then what might be said about the proposed functionally existent self? For one, it must be acknowledgedly personal, and as personal it incorporates that subjective outlook (qualia, ‘what it is like’) with which we are all abundantly familiar but which seems exceedingly difficult to standardize; even within our own species we find ourselves in trouble when we try, as the illustration of the sky’s hue viewed by a color-blind person pointed to, or our earlier thoughts on individuated interactions with a rose. As far as the self is concerned, this intimately known phenomenology is likely as close as we can get to an empiricist’s insistence on ‘nothing but the facts’, but as the preceding attempted to show that requirement is anyway misplaced or misapplied. The self-view that I hold is critical for the life I lead because it determines how I understand myself and what I therefore am and become: my traits, character, emotions, hopes, dreams, ambitions, actions, successes, failures, good fortune and bad, inclusive too of the world that I inhabit – Sets 1, 2, and 3; background thoughts and preconceptions here determine the whole of one’s view, the very world that one finds displayed before one, comprehension is always intertwined with interpretation – and in that decisive operation this functional self seeing/self generating gains whatever substance it can be thought to have through the lived noticing and conveying of it, minute by minute and second by second, from cradle to the grave. Unless we are philosophers, Buddhists, or other particularly reflective sorts, we typically take this functioning self to be real and objective in the traditional or Cartesian sense due to the articulated role that we have been discussing, and for all intents and purposes it is indeed real and objective enough. Perhaps it is not objective in a strictly absolute sense (as anti-realists have been at pains to

demonstrate), but then what is? We may add: What in this universe of incessant change and temporality is objective in any ultimate sense? The soft realist self's reality is its functional existence: real because active, substantive because expressed.

There remains a point of worry in the self model as presented, in our self as real as a causal agent abstracted, real as a directional emergent property on the conceptual level. We defined the self as Set 1 + Sets 2 & 3 + C + B_p, which is a fairly weighty and detailed definition, containing four (or five, if the straightforwardly emotional Sets 2 and 3 are taken separately) distinct elements all of which must be present to render the summative self. Of these, bodily presence (B_p) is the least problematic as its absence will simply mean death,⁴³³ and as regards consciousness (C) previous chapters put forward an argument for considering it to be continually present while alive (even in cases of deep sleep or other instances of reduced awareness), but what of the three emotional sets? If they, or any one of them, are missing would that not mean that the self too is not only incomplete but, by the additive nature of the definition, incompletable? In attempting to answer this challenge a review of the proposed contents of the three sets will prove helpful.

Set 1 is composed of the self-constituting 'emotions', and 'emotions' is placed in single quotes/inverted commas because in Kristjánsson's work, which provided the starting point for our own study, the items involved are labeled as emotional although they are really more identitarian – and so conceptual – in nature; Kristjánsson perchance chose to do this merely in order to simplify his classification as his other two sets are both unambiguously emotional in content.⁴³⁴ It contains things like one's preferences, outlooks, genetic

⁴³³ The reader will recall that we earlier considered and rejected cases such as those entailing an envatted brain or the like; see Chapter 2 Section 2.3 ('Bodily and embodied issues').

⁴³⁴ See again Kristjánsson *op. cit.*, and in the present Chapter 1 Section 1.1.C ('Four accounts of the self: The soft realist position'), and Chapter 2 Section 2.2.A ('Psychological issues and the self: Cognitive structure').

inheritance, upbringing, considered choices such as goals, aspirations, et cetera. Sets 2 and 3, on the other hand, are one's self-comparative and self-conscious emotions, respectively, and they entail the affective and intuitive reactions which continually surveil and respond to the whole self in its lived environment, and in doing so serve to maintain, shift, sustain, and/or adjust it. These three sets naturally interact with each other in a formative manner and such will be understood to occur mostly beneath the surface of awareness, although one can also (and most of us probably regularly do) reflect on one's self in the sense of thinking about the type of person one is and/or the behavior one exhibits, or would like to become and/or behave; such a self examination would fall too into the domains of Sets 2 and 3. The reader will remember that the emotional sets' preconscious interplay would be a part of the brain's System 1 functioning (intuitive, fast, automatic, effortless), while the purposely reflective type is System 2 (rational, slow, determined, laborious).⁴³⁵ The trouble then comes down to this: If one is in a state of deep sleep or anesthetized, are these emotional sets still being psychologically supported? Further, if they are not, then can the self on our definition of it still be said to exist? To be real?

I must admit that I do not know what happens to our emotional sets in such states. At a glance it does seem entirely possible that they continue, for when we come out of those conditions we usually have lost none of the essential components involved, and indeed if one tried to one could instantly (rationally) engage or detect their working; but how could one determine whether or not the three sets continued in any way while awareness was gone even if consciousness – in the pre- or non-aware way discussed in Chapter 5 – itself was ongoing? Would it be possible to devise a test that recorded such? One idea might be to monitor a sleeping subject and then suddenly awaken them and ask a series of questions

⁴³⁵ See Kahneman 2003 and 2011, *op. cit.*; Kahneman's works are considered here primarily in Chapter 2 Section 2.5 ('Certain uncertainty, randomness, and limited choice').

about their preferences, goals, et cetera, but surely those inquiries would be easily answered under normal circumstances, and even in abnormal circumstances (e.g. coming to after a mental event such as a seizure) following a short amount of time the memory recovers and the appropriate responses can be made. In instances of a stroke or the kind of schismatic break that Friedrich Nietzsche underwent towards the end of his life⁴³⁶ the situation is less clear but it still seems as if something at least of Set 1 would remain, and as for Sets 2 and 3 their roles are primarily of an auditing and responding nature and so regardless of the condition of what was being engaged with the sets would still function whenever the subject were aware. Yet we are now wondering about those moments when the subject is not aware, when still in the midst of their deep slumber and long before we have woken and pestered them with our queries. What then *during that time*?

I am tempted to think initially that Sets 2 and 3, as self-comparative and self-conscious emotions, would be nowhere, but then I remember the preconscious mode of the mind (System 1), with its automatic and unceasing intuitions, judgments, and – yes – emotional responses, and my confusion deepens. The argument could be made for instance that Nietzsche, post the incident in Turin, was such a different person that he was no longer Nietzsche in the sense of the self he had theretofore been (and that argument could work on our model as we have situated personhood at a level above selfhood), but that much seems obvious and what we are trying to determine is not about that, it is about what happened to Nietzsche's self when he laid down for the night – before or after Turin, it makes no

⁴³⁶ The reference is to the onset of Nietzsche's madness following his public breakdown in the Piazza Carlo Alberto in Turin on 03 January 1889, the exact cause of which remains unknown but the debilitating effects of late-stage syphilis has often been proposed, e.g. the philosopher, translator, and writer Walter Kaufmann lists the source as an 'atypical general paresis' brought on by syphilis; see the 'Introduction' to *The Portable Nietzsche*, ed., trans., and intro. by Walter Kaufmann (New York: Viking Penguin Inc., 1954), p. 13; for an overview see also 'Friedrich Nietzsche: Psychological illness and death (1889-1900)', *Wikipedia: The Free Encyclopedia*.
<https://en.wikipedia.org/wiki/Friedrich_Nietzsche#Psychological_illness_and_death_.281889.E2.80.931900.29>.

difference. In reasoning out all this I want to say that if the brain is still working, and especially considering that a degree of consciousness is presumably always maintained, then Sets 2 and 3 are present in their normal preconscious modes, and further that since the materials of Set 1 are (typically) instantly recallable when awareness returns they too linger in some way all of the time,⁴³⁷ but I do not know how to make a presentation for these conclusions that would be convincing enough to handle all of the many ‘what ifs’ that could be thrown into the mix. Although perhaps in the end it does not matter all that much since the realness of our self is not of the hard variety, and since too our definition not only admits of self-change it insists on such as a mandatory condition. Our self is a self that is always in motion, and its reality comes from its function; is it not therefore enough to simply remain on the point that if a brain is functioning sufficiently to operate the three emotional sets in a self then a self will be generated (or sustained, et cetera) and be real in the soft manner described? The doubters will likely not be satisfied by this, but again I have no means by which to placate everyone, and when it comes to a topic like the self sooner or later one just has to take a stand. Maybe this is where I do so.

6.2 Time, the made self, and the making self

We have proposed a self-view based on the emotions, consciousness, and the physical presence of a body, and we have claimed a soft realism for it which encompasses the condition of being permanently psychologically held but not static. The self concept outlined above is just that: a concept, an idea, an abstracted internally given *thus* that gains its realist weight through the guiding, informing, structuring, and animating features that it bestows on the individual,⁴³⁸ who both has and is that self, and who adds to and takes from

⁴³⁷ I refer here to such Set 1 elements as goals, aspirations, preferences, and the like, other Set 1 factors such as genetic inheritance or one’s upbringing would naturally persist.

⁴³⁸ I.e. whole person, which on our layered definition consists of embeddedness plus personal identity, with personal identity in turn made up of the self plus contingent facts about the body, other contingent facts, and feedback from the social realm. See again Chapter 2 Section 2.2.D (‘Psychological issues and the self: Selves,

this core throughout the course of his life. Any attempts to physically pinpoint such a self would be nonsensical, yet so too would be the wholesale denial of the importance such plays in how each of us views our own natures and positions in the world – and it is out of those views, out of those self understandings, that we move and grow, struggle and strive, hope and wish, despair and deny, yearn, reject, accept; all of the qualities that make up life for us on this planet at this time. We are speaking here of the experientially real, not the empirically real, yet we nevertheless – it has been argued – are still speaking of the real.

Our study has for the most part examined *what* this self is, its contents and make-up, and out of that its foundational role, and so now to conclude I would like to shift our focus to *how* this self is: its maintenance, adjustments, transformations, its ever-becoming. For this I will cover three aspects related to the personal experience of time, centrally considering three primary researchers on the topic, namely Husserl, Heidegger, and Henri Bergson. The three aspects that I wish to entertain are: 1) that of the unexamined or default position, which I will argue leads to the ‘made’ self, 2) that of the purposive and aware position, which I will argue leads to the ‘making’ self, and 3) that of a phenomenologically-centered theoretical and transcendental position, which is not directly based on the aforementioned thinkers, and which I will argue leads to a view of time that is not linear but spiraling within the confines of a human life, and that this ‘shape’ is due to the important element of the meanings that we assign, a factor that can significantly contribute to the ‘making’ self.

As might be expected, most of the work done on time has been concerned with our daily, passively lived experience of it, which is to say with our ‘there’s never enough time’ or ‘just idling away the time’ or ‘time flies’ type attitudes. Husserl points out that whenever we take note of the passage of time there is necessarily an instance of memory involved,

personal identity, and whole persons’).

that ‘it is consciousness of *what has just been* and not merely consciousness of the now-point’.⁴³⁹ Moreover, this now-point is simply an ‘ideal limit’ and not anything in itself, rather it continually interacts with the not-now and hence the now is always moving from perception into memory,⁴⁴⁰ so that our personal experiences of time are ever subjective (and absolutely so), based on the now and what we consciously (with or summonable into awareness) retain. Husserl writes that we therefore cannot even approach objective time (which, he also claims, does not change), bound as we are to our observational flows with their constant referencing back.⁴⁴¹ Living in and thinking about the world in this manner – as we instinctively do, and in many ways are required to do given everyday concerns – lends great weight to the pragmatic and to our ongoing must-be-dealt-withs, so much so that Husserl states that this perception and relation gives rise to what we might call a ‘face value’ notion of self, that:

what suffices as practical counts as the Self. Thus the house itself and in its true being, and specifically with respect to its pure bodily thingly nature, is quickly given optimally, i.e. experienced as complete for that person who regards it as a buyer or a seller. For the physicist and the chemist, such ways of experience would seem completely superficial and miles away from its true being.⁴⁴²

Here is our first hint at what I am labeling the ‘made’ self, the default, unintentioned, and non-willed self that results merely from being alive as the type of creature that we are. By the terms of our account, in practice this amounts to taking one’s three self-formative

⁴³⁹ Husserl 1999, *op. cit.*, p. 191; emphasis in the original; taken from Husserl’s *On the Phenomenology of the Consciousness of Internal Time*, Section 12.

⁴⁴⁰ *ibid.*, p. 196; from *On the Phenomenology of the Consciousness of Internal Time*, Section 16. Here Husserl is referring to his notion of ‘primary memory’, which he distinguishes from ‘retention’ in that the first is directly connected with a percept while the second records and follows a flow.

⁴⁴¹ *ibid.*, p. 209; again from *On the Phenomenology of the Consciousness of Internal Time*, Section 35-36.

⁴⁴² *ibid.*, p. 233. This is taken from *Analyses Concerning Passive and Active Synthesis: Lectures on Transcendental Logic*, Section 4.

emotional sets and doing absolutely nothing with them. Using the above example, this is the consideration of oneself (i.e. whole person, not of one's *self*; there is no intentional self directed reflection here) in the same way that one views the house as a seller or a buyer. One might still have acknowledged or unacknowledged desires (a part of Set 1) and might make efforts to achieve them (tracking such with Sets 2 and 3), but one will not thereby be actively working on one's self in its core definitional sense on our model; rather the entirety will be externally focused and more than likely in the sense of attainment ('getting that'). Note also the close connection Husserl is drawing between time as a sensed flow joined to memory and the notion of a 'practical self', a comprehension which takes the surface for the whole. (As Husserl puts it, the physicist and chemist know better; as we might put it, an intentionally aware and self reflective person would know better – but that can wait for our second aspect regarding time.)

In broad measure Heidegger follows his former teacher in his placement of experiential time from the 'standard' (unreflective) perspective, although he greatly expands its place and its scope. As with Husserl's focus on memory, Heidegger takes the having-been as the grounding for our moods, even for future oriented moods like hope, for in hope (and other similar frames) an individual still operates from a having-been and towards that which is hoped for. Heidegger does however call attention to the circumstance that this is only the case for 'attuned' creatures who exist in an always having-been, and that some nonhuman animals may more probably exist purely in a 'now'. He adds that how time is constituted for such organisms is a separate problem,⁴⁴³ but his raising of the issue is helpful in that it brings our focus onto the role that our own consciousness structures play in the manner in which we experience time, and therefore too in what we theoretically make of it. What time might 'be' in a universal absolutist sense is beyond the scope of the present study, but

⁴⁴³ Heidegger 2010, *op. cit.*

what it might ‘be’ (or ‘mean’) for us will be considered further in the below, especially in the discussion on our third listed aspect; for the moment I offer that perhaps there ‘is’ only the now of the nonhuman animal, although for us human animals, given our mental engineering, the now is always felt as a was-and-is, and our thoughts are (almost) always on the what-will-be. This is naturally a fairly trivial observation, but it is nevertheless important to bear in mind in a study on the self, particularly if we wish to think about the self’s made and making functions.

Returning to Heidegger while staying on the present (now), he writes that for Dasein the present ‘arises from’ the future and the having-been, that it involves ‘being entangled in lostness’ and ‘swept along’ in our thrown (enworlded⁴⁴⁴) condition, and that the now moment can attain no horizon of its own without an act of resolution on the part of Dasein that acknowledges the conditions of both the situation and of its own ‘being-toward-death’.⁴⁴⁵ The future, past, and present are in fact forever so deeply connected for Dasein, Heidegger insists, that ‘The future is *not later* than the having-been, and the having-been is *not earlier* than the present. Temporality temporalizes itself as a future that makes present, in the process of having-been.’⁴⁴⁶ This detail of our experience, however, is a relatively closed one when we operate in the typical ways that we do, for we tend to overlook the ‘place’ (within the embedded contextualization) of that which we encounter, and because too of the manner in which our interpretations are based on acquired concepts so that our outlooks and understanding are limited, Heidegger informs us, in ‘the relevant possibility of truth and certainty, the kind of grounding and proof, the mode of being binding and the kind of communication – all these will be determined.’⁴⁴⁷

⁴⁴⁴ That is, embedded and world-forming (inclusive of ideas such as those expressed by Husserl’s use of ‘horizon’; see Chapter 1 Sect. 1.2.D (‘Four accounts of the self: The contextualized soft realist position’)).

⁴⁴⁵ Heidegger 2010, *op. cit.*, see pp. 332-333 (Section 68(c) in Heidegger’s original manuscript).

⁴⁴⁶ *ibid.*, p. 334 (Section 68(d) in the original); emphases in the original.

⁴⁴⁷ *ibid.*, p. 345 (Section 69(b) in the original).

These thoughts echo those of ours above in the first section, where we considered how background and preconceived thoughts impose themselves on the manner by which one's world is viewed, shaping and guiding, shutting out this while making room for that. This indeed is the partially self-forming element of limited choice that we have considered throughout our study.⁴⁴⁸ Heidegger puts the case thusly, 'Insofar as Dasein temporalizes itself, a world *is*, too... The world is neither objectively present nor at hand, but temporalizes itself in temporality... If no Dasein exists, no world is "there" either.'⁴⁴⁹ The use of quotation marks/inverted double commas here is crucial: this is an ontological statement of meaning for the situated self and not an empirical assertion. The reader will remember that Heidegger's 'world' is not the Earth as such, it is the completeness of that in which an individual (Dasein) dwells, the entirety of their embeddedness and their approach to and discernment of said entirety. These thoughts take us to the last portion we will look at of Heidegger's analysis of experiential time as taken from the commonplace existential position (the 'made' self): its publicness.

Due to the way in which we live, our 'just passing through life' as Heidegger (rightly, I think) puts it, we do not consider our time to be a pure string of nows – we do not notice it as such, or perhaps more accurately do not bother to take note of it as such – and so experientially time becomes 'covered over' (presumably Heidegger means by activities, by busyness, by the suchnesses of I-want-a-bite-to-eat and oh-no-my-train-is-coming, et cetera), and the result is that time for us is filled with 'gaps'. In its 'world'-ed (enworlded) condition, Dasein does not 'take' time for itself but instead 'heedfully *exploits* the time that "there is", the time with which the *they* reckons.'⁴⁵⁰ It is the fact of thrownness, of being unreflectively/non-noticingly in a world, that accounts for this feature of time as public, as

⁴⁴⁸ See especially Chapter 2 Section 2.5 ('Certain uncertainty, randomness, and limited choice').

⁴⁴⁹ Heidegger 2010, *op. cit.*, p. 348 (Section 69(c) in the original); emphasis in the original.

⁴⁵⁰ *ibid.*, p. 391 (Section 79 in the original); emphases in the original.

shared and therefore reckoned together. Time is social Heidegger asserts, or at least socially bound, and if so we may begin to wonder if there can be such a thing as an objective time. Husserl, as mentioned above, stated that there is objective time, that it does not change, and that in our limitedness we cannot get near to it. Heidegger, as will be expounded on in the below on our third aspect, contends that there is not, but also that time is equally neither subjective. Bergson, whom we will now consider, has an even more creative approach.

Bergson begins by differentiating number and the concept 'number' as an example of how we tend to think in spatial terms (e.g. thinking about numbers by mentally placing quantities of objects side by side), and then moves into a discourse wherein time is equated with space on the grounds that time, in reflective consciousness, is a medium that allows our discrete conscious (i.e. aware) moments to be viewed as a series, and therefore to be countable; coupled with the manner in which the concept 'number' involves a spreading out in space (counting), time can thereby be understood to actually *be* space. Bergson writes that:

we set our states of consciousness side by side in such a way as to perceive them simultaneously, no longer in one another, but alongside one another; in a word, we project time into space, we express duration in terms of extensity, and succession thus takes the form of a continuous line or chain, the parts of which touch without penetrating one another.⁴⁵¹

I think that even on these terms there is however room within this conception for a

⁴⁵¹ Henri Bergson, *Time and Free Will: An Essay on the Immediate Data of Consciousness*, 3rd edn, trans. by F. L. Pogson (Mineola, NY: Dover, 2001 unabridged republication/London: George Allen and Co., Ltd., 1913), p. 101; note Bergson's traditional use of 'conscious' and 'consciousness' as equating 'aware' and 'awareness' on our terms.

non-strict equivalence, or at least arguably so. Bergson states that time is experienced by us in this way only because we endure, but that if we were to withdraw the thinking ego there would be no succession but only the moment and its X, and conversely if we take away the X there will not be anything but the ego ‘without moments external to one another, without relation to number.’⁴⁵² These musings are reminiscent of Descartes’ doubt process, and if we follow the game and imagine a timelessly unaffected object X against a timefully affected ego in this way, we find, or at least I find, a number of questions asserting themselves. First of all, if Bergson’s claims really are so then can time actually be said to ‘be’ space? Can time be said to ‘be’ anything other than the method by which we relate to and orient our lives? A succession-less ‘time’ and a dissociated ego absent any impression of moments looks very much like the disappearance of time in a maintained space. Does time, therefore, *exist* in any form other than the semi-arbitrary arrangements we have devised for measuring it? Put another way, does time have any meaning for the self at its core level, or only for the whole (enworlded) person? The sun goes up and comes down and we count that as a day, subdividing it into portions of seconds, minutes, and hours based on units of sixty – but sixty has been chosen for historical reasons (a legacy of the Sumerians and Babylonians⁴⁵³), and empirically speaking the sun does not after all ‘go up’ nor does it ‘come down’, in fact it does not budge at all except in our abstracted figurations. The Earth moves and would still move were we not here to count its rotations; would time or ‘time’ then remain – in the concrete or in the conceptual – as anything substantive without any creatures around who obsessively made notations about it? The case is unclear to say the least, and perhaps this helps explain why Husserl could label time as objectively existent but unapproachable while Heidegger judged it as both non-objective *and*

⁴⁵² *ibid.*, p. 108.

⁴⁵³ Michael A. Lombardi, ‘Why is a minute divided into 60 seconds, an hour into 60 minutes, yet there are only 24 hours in a day?’, *Scientific American*.
<<https://www.scientificamerican.com/article/experts-time-division-days-hours-minutes/>>.

non-subjective. Let us though continue our look at Bergson, with the reminder that we are here not trying to settle any ontological issues regarding time but rather investigate its application to the self.

To try and forestall any misunderstandings let me state outright that the preceding questions (and those that follow) are not intended to discount the weight that our approaches to time take in our lives – rather the opposite – for surely it is the case that how we consider time in the way we speak and think extends its influence far into the shaping of the self (particularly through Set 1’s elements like goals, aspirations, plans, objectives, desires, wishes, et cetera). Yet regarding Bergson’s argument, if definitionally time without the ego (‘ego’ in the psychological sense as per Bergson’s use above) can be boiled down to this ‘only the moment’ and X, as he directed us to think, then – again – need we even say that time ‘is’ space? Why not simply put it as time is not, or time is only the perceived now? That would seem to attune time more closely with the self and its involvements. The answers to all of these questions, of course, will depend on whether we are taking ‘time’ in an ultimate sense or in its meaning of ‘duration’, and it is in that latter nuance where Bergson considers time to really hold experiential relevance for us; we therefore may find a deeper pertinence to our purposes taken from that angle. Let us carry on.

Bergson brings us back to the importance of our mental modes in the applicability of time, showing that succession and duration only exist for us due to the ‘interpenetration of our conscious states’ and that we must not be led astray by such linguistic devices as ‘between now and then’ to think that anything more is going on than the mind noting a number of simultaneities. Intuitively and naturally (and, says Bergson, ‘probably what animals perceive’), time is experienced as a quality, and it only becomes an issue of quantity when

set in (that is, projected into) space.⁴⁵⁴

On this broader issue of quality and quantity, Bergson explains in a later work how the Ancients (in the Western tradition) did not make the same existential (categorical) distinctions that we do today, and that for them the barriers between the aspects of body and mind were of a far different and much more permeable order. Mathematical (number) concepts were like any other concept and fit into the overall hierarchies of Ideas, while the body was not purely geometrical and the soul/mind not purely consciousness; instead the psyche contained the same vital force of a living body and the body was less corporeal, itself too containing the Idea.⁴⁵⁵ Being blurs once more.

I should highlight, by way of explanation or caveat, that in relating this last portion of Bergson's analysis I do not mean to make a case for a neo-Aristotelean or a neo-Platonic view of the cosmos, rather I simply wish to point out that yet again, as in our discussion of consciousness in Chapter 5, we find that the way and the level upon which we approach a topic has ramifications for our understanding that are far greater and deeper than we may at first suspect. How we talk can mask layers and layers of background assumptions, and in so doing predetermine our conclusions and blind us to options that would otherwise hold great appeal.

Taken together the overall picture that emerges from our study of our first listed aspect is that in the standard and everyday functioning the approach we have to time is to view it as an external element whose forces act upon us and pressure our behaviors and decisions. Time is 'out there' and because it is so I must X very carefully in order to Y two hours

⁴⁵⁴ Bergson, *ibid.*

⁴⁵⁵ Henri Bergson, *Creative Evolution*, trans. by Arthur Mitchell (Mineola, NY: Dover, 1998 unabridged republication/New York: Henry Holt and Co., 1911).

from now so that Z will be at all possible, et cetera, et cetera. From a practical point of view this is not strictly incorrect, and living in the societies and having the demands that we do it even makes good pragmatic sense to comprehend time in this fashion. Yet to do so does nevertheless shift one's focus from a directing to a directed mode – or rather it maintains the default outlook of the directed –, the passivity of the receiver as against the activity of the giver. Accepting time in this way keeps one stuck in the neglected self, in the 'made' self of the contextual (embedded) forces which undoubtedly do make up *a part* of who we are (a part of the selves we have and are, via Set 1), but need not be the whole of the story. Far more possibilities open for us to the degree that we strive instead to create our selves through a purposive and intentional control (to the degree possible) of the pathways of our becoming. The manner in which we perceive time is an important element of this process.

If we continue to believe (or merely accept) that time is something extrinsic, that it is powerful and objective, then we will be subjected *to it* and our selves will have a greater amount of receptivity and be less responsive to whatever our goals for transformation or transition might be (assuming that we have at least some goals). In short, we – our selves – will be buffeted by the winds and waves of chance, by the circumstantial, the situational. What control might be available will not be taken advantage of, and we will float steadily along to death in a life marked by half-measures. For obvious reasons this 'made' self, for all of its commonality in the hyper-modern world we occupy which scarcely allows one a chance to catch one's breath, is what the wise in our midst have consistently argued against. What has conversely been argued for is the 'making' self, the directed and active – the directional – self, but to get there, I think, we need to see time differently. This will be the second aspect that we will explore.

To approach time in a purposive and aware manner, to take it as something internal rather than external, runs counter to the standard understanding as described above, but I believe it to be an important first step on the way to self-making. In the following we will again return to the same trio of thinkers for a theoretical foothold to assist us in this launch – or lurch – into an alternative time, and thereafter we will apply such thoughts to a creating self. To begin, Husserl, with his focus on memory, astutely teaches that remembering is not apprehension, not perception, and that what is held in memory is fundamentally different from what activated our senses;⁴⁵⁶ a condition which can perhaps be structurally compared to an awareness and having an awareness of awareness. To garner such (the latter) implies a certain degree of mindfulness, a step back from the rush of the everyday, a meta-point of view, and it is easy to recognize how seldom this happens in life and how much effort it requires. Regarding the perception of objects, Husserl stresses the impossibility of any external view ‘exhaust[ing] the sensuous-material content of its perceived object’ or of presenting to the beholder/experiencer something ‘literally from all sides at once in a self-contained perception.’ Instead we have access at any given instance only to an ‘appearance core’ upon which we impose ‘a system of referential implications’.⁴⁵⁷

This whole process must be natural and it must be automatic, as our study has stressed in its consideration of the systems and layers upon which our mental mechanisms are run. It should also be kept in mind that in this analysis the word ‘perception’ does not merely indicate the visual, and it may not indicate sight at all; instead it is an issue of comprehension, and with that all of the attendant conceptual blinders and guides, fences and walls, that hinder each one of us from ever truly participating in the wholeness of the realities that we encounter. Out of these limiting factors one’s horizon looms, one’s

⁴⁵⁶ Husserl 1999, *op. cit.*; from *On the Phenomenology of the Consciousness of Internal Time*, Section 12.

⁴⁵⁷ *ibid.*, p. 221-222. This is taken from *Analyses Concerning Passive and Active Synthesis: Lectures on Transcendental Logic*, Section 1.

possibilities become defined, and one's world – in the Heideggerean sense – is created.

Husserl describes the situation thusly:

everything that genuinely appears is an appearing thing only by virtue of being intertwined and permeated with an intentional empty horizon, that is, by virtue of being surrounded by a halo of emptiness with respect to appearance. It is an emptiness that is not a nothingness, but an emptiness to be filled-out; it is a determinable indeterminacy. For the intentional horizon cannot be filled out in just any manner; it is a horizon of consciousness that itself has the fundamental trait of consciousness as the consciousness of something.⁴⁵⁸

Here we have a clear picture of what is going on in the undertaking of a 'making' self, of what is involved in an intentional and willed built/created/formed/directed self. Time's great slate – approached as an ever-now – confronts the whole person seeking to make their self, but it must be understood that such cannot occur out of a pure 'nothing', a void, and nor can any current trajectory be entirely changed. Any individual's now has its history, and the scaffolding of self has already been laid down by those forces that we may label as historical chance (including personal biological, cultural, familial, economic, et cetera, ingredients) or that we may determine to be fate. However we choose to consider the epochal details we have inherited, they are there and it is out of them that we are able to begin to dirty our hands on the potter's wheel, giving shape to the self we would become. This is an incremental *discovering* and constructing, and it can only be done if fully aware in the now.

Heidegger too emphasizes the importance of the factors out of which we attempt to gain

⁴⁵⁸ Husserl 1999, *op. cit.*, p. 223; taken from *Analyses Concerning Passive and Active Synthesis: Lectures on Transcendental Logic*, Section 1; the reader is reminded again of the equivalent usage of 'consciousness' here with our 'awareness'.

the perspective on our selves and our existence that we need if we are to succeed in any ‘making’, writing that ‘If the thematization [i.e. projection of understanding, interpretation] of what is present...is to become possible, *Dasein must transcend* the beings thematized’; yet if this thematization comes out of one’s concern for one’s self, place, and activities in the world, then ‘a transcendence of *Dasein* must already underlie “practical” being together with things at hand.’⁴⁵⁹ We are able to take this step back into purposive awareness because we are really already there, the complexities of the consciousness systems (or network) we have both enable and enact it: we are not bound mentally and conceptually/perceptively to only the environmentally given as many nonhuman animals are. Time for us need not strictly be the ‘time’ of ‘space and time’ but can instead be the ‘time’ of absorption and targeted being. To engage with time in this sense, I think, is to abandon the tyranny of *chronos* that is the hallmark of the default position (characterized in the ‘made’ self) and instead to embrace time as definable and not as defining, as the ‘what I do with it’ and not as the sweeping hands of the clock. It is from this type of point of view that one’s self can be purposefully formed. I do not mean to suggest that this attitude entails a denial of time in its public meaning (as per Heidegger’s labeling of it) for the practical naturally cannot be ignored; rather that this thinking involves a personal commitment to undergoing the desired self transformation through an application of already existing abilities and a judicious use of available resources.

Bergson makes some similar points with regards to awareness and the comprehension of time in an ultimate sense. In fact, he finds the idea of time notionally abstracted into a force of some manner to clearly be a false one. It will be remembered that on Bergson’s analysis time, when thought of as a given and undifferentiated whole, became equated with space, leading him to further conclude that if viewed as a pure medium of some kind time

⁴⁵⁹ Heidegger 2010, *op. cit.*, p. 346 (Section 69(b) in the original); emphasis in the original.

as a concept is ‘spurious’ since it can be fully reduced to space and there cannot be two forms of the same medium (time and space).⁴⁶⁰ On this – for my part – I concur, as was indicated in my list of questions and responses given above. Time as duration, however, is another matter. Bergson writes that, ‘Pure duration is the form which the succession of our conscious states assumes when our ego lets itself *live*, when it refrains from separating its present state from its former states.’⁴⁶¹ We can agree with the latter here whatever our thoughts on time per se as space might be, taking Bergson’s ‘pure duration’ as an awareness of the now as it happens, as a living in the ever-now that maintains an observation on one’s self as it develops in its becoming.

In Bergson’s later thought, and in regards to a scope of focus far beyond ours here but nevertheless germane, he places becoming in a central role, interestingly reversing the Classical view of Forms degenerating into matter and becoming by arguing that if the Forms, as concepts, are seen to be ‘snapshots’ of an always changing reality then for us they are simply moments gathered. No longer generative, they are instead ‘artificial reconstruction[s]’ and ‘symbolical expression[s]’.⁴⁶² With becoming at the core of a ceaseless transitioning we are able to understand duration as creation (i.e. as the constantly being made/re-made), and Bergson extends his point with ‘if that which is being unmade endures, it can only be because it is inseparably bound to what is making itself.’⁴⁶³ As for the universe, we might say, so for the self. Bergson challenges us to see becoming, and not degeneration, as the driving force behind the macro-cosmos – again, maybe it is, maybe it is not, this is certainly not the place to enter that discussion –, and if we can at least

⁴⁶⁰ Bergson 2001/1913, *op. cit.*

⁴⁶¹ *ibid.*, p. 100.

⁴⁶² Bergson 1998/1911, *op. cit.*; see especially pp. 315-318. Note that the original French version of this book was published in 1907 while his *Time and Free Will* appeared in 1889. Bergson’s bibliography can be found on his *Wikipedia* page here: ‘Henri Bergson’, *Wikipedia: The Free Encyclopedia*. <https://en.wikipedia.org/wiki/Henri_Bergson>.

⁴⁶³ *ibid.*, p. 343.

entertain the idea and so adopt (to some extent) its conceptual implications and ramifications then we can shift our thinking regarding the micro-cosmos of the self to allow for its ‘making’ potential. In this we are thereby able to start to *be* as we would and not simply as we ‘are’, and we start too to ‘take’ time and not simply to ‘watch’ it.

In our third and final aspect on time we will go deeper into the theoretical and consider how phenomenologically time, bound up with meaning for us, appears to move not in a straight line but in a series of sometimes expanding, sometimes shrinking, spirals. The thoughts we will propose here are not found in the works we have so far examined in this section, but there are some hints at a more transcendental view of time within them. Heidegger, as mentioned, states that time is neither objectively nor subjectively held, but is the possibility for both, or more precisely the possibility of the possibility of both. He writes, “‘Time’ is present neither in the ‘subject’ nor in the ‘object,’ neither ‘inside’ nor ‘outside,’ and it ‘is’ ‘*prior*’ to every subjectivity and objectivity, because it presents the condition of the very possibility of this ‘prior.’”⁴⁶⁴ *Being and Time*’s closing question is also relevant here, and I feel too that it is explanatory of Heidegger’s overall position: ‘Does *time* itself reveal itself as the horizon of *being*?’⁴⁶⁵ This is to ask whether or not being is bound by time, but I think rather it is the reverse; allow me to make this case via one more look at some ideas of Bergson’s that may juxtapose with the metaphysical analyses conducted in the opening section of this chapter.

In the context of arguing against the idea of a void or pure nothingness, Bergson situates time as a concept squarely in the present, bringing our attention to the fact that for a creature without memory or an ability to project a future (‘prevision’ in his wording), terms

⁴⁶⁴ Heidegger 2010, *op. cit.*, p. 399 (Section 80 in the original); emphasis in the original.

⁴⁶⁵ *ibid.*, p. 415 (Section 83 in the original); emphases in the original.

such as ‘void’ or ‘naught’ would not even be used, only the perceived would be spoken of and that is necessarily a presence, not an absence. For us who have a memory and who make use of a capability of thinking about the future, when we expect to find something but do not we might linguistically employ ‘naught’ or ‘nothing’ or ‘void’ or the like, but when so, Bergson insists, the instances are more akin to feelings than thoughts, with our speaking ‘tinged’ with the failure to be realized. He goes on to argue that to think an object nonexistent is first to think it, and therefore to think it existent, and then to *add to* that the idea of its exclusion. Bergson puts it this way, ‘The act by which we declare an object unreal therefore posits the existence of the real in general.’⁴⁶⁶

What is the real in time? In the first section above we argued for a (type of) reality based on function and effect, labeling our self-view understood that way as being ‘functionally existent’; could the same or similar thoughts apply to our experience of time? Time and/or the idea of time certainly impacts our lives, and as we explored earlier a part of the default or unreflective attitude towards time is to assign it a great importance. We have however sought to establish a distinction between ‘time’ in a physical sciences sense and ‘time’ in a phenomenological sense – with the latter being akin to what Bergson called ‘pure duration’ – and then to apply that to the ‘making’ self, the purposive and intended self that can become an act of creation for the whole person who through awareness and will takes such on. If the self is a project in this way then ‘I am my time’ is a statement of some literalism, for how I choose to spend the time I have (i.e. physical sciences or ‘public’ time) is what fuels the ‘me’ in the undertaking. I am my time – ‘I’ am my time – I ‘am’ my time – I am ‘my time’; however we emphasize it the end product is the same: the manner in which whole person ‘I’ arranges existence results in the formational and forming elements of core self, which in turn feed back into personal identity and the entire relational chain of whole

⁴⁶⁶ Bergson 1998/1911, *op. cit.*; see pp. 281-285, quoted section from p. 285.

person. My 'I' or 'me' is expressive of and founded upon the practice in which time is allocated, and awake or asleep makes no difference on our view of the self for arguably all conscious and preconscious – including the non-aware – states contribute to one, some, or each of the definitional elements of the self. This points at a heavy burden for those who would live actively enough to take charge of their self in the 'making' sense that we have used, and hence a move away from how time is considered in the everyday 'made' self position could be of great advantage. Yet is such a move possible?

I believe it is, and I think the basis for it has already been found in the case we have made for one's preconceptions and ideational foundations (to a large degree) determining the way in which one takes in the world one inhabits, that there is no purely 'objective' view from which one can escape an inevitable perspectivism. This conclusion seems particularly applicable when we take sociohistorical influences into account; picking up almost any text reminds the reader of just how deeply we speak out of the milieu of the years we happen to live in, and of how the reach of that which surrounds us and (especially in our current media-driven era) relentlessly informs us plays into our interpretative stances. Taking the informed decision to regard time differently than the purely *chronos* version is therefore not only an option but it is a legitimate option and not a case of self-deception, an outcome strengthened still further when we focus on the experiential – as we are attempting to do here –, for to me as an individual living in the world and in the midst of a constant barrage of stimuli and the resulting analyses, apprehensions, judgments, choices, et cetera, there are really only two versions of time: Now and Fantasy. When I imagine the future or recall the past I am transmitting out of a now, and as Bergson highlights I also do so in an imperfect way. Exercises of this sort will affect the self I am and am becoming by providing internal feedback and emotions based on self-comparisons and self-consciousness, but all such envisages are informative only of the now and yield their

influences only in the now. Speculating on what may be and remembering what (I think) was are games that the mind plays in the moment. The ever-now is what is experienced, it is solely what is experienced, and it is the whole of the *phenomenology* of time for the kind of animals that we are, even if it is not the whole of the *conceptuality* of time that we may or may not have. Such is moreover not reliant on the content of the mind's now (whether thinking about past, present, or future is irrelevant – I am thinking *now*); it is only in the eternal present where the self takes shape, and if that self is to be a 'making' self instead of a 'made' self then administering to the present and one's own existence in it – a purposive focus on all of the constitutional elements of the self – is a matter of tremendous *apropos*. When we are able to extricate our awareness from the busyness of modernity and actually accomplish this, we facilitate and empower our self formation and turn our lives into creational works. It is in this sense that time becomes a matter of being and not an outside juggernaut crushing us from above as it passes.

In staying with the phenomenology of time as seen from an individual's point of view, what is it that we find when we look at our own pasts? However deteriorated or robust our memories may be, we discover primarily that particular events and ages (i.e. when I was X years-old) stand out far more clearly for us than others, and do so because of the meaning we have assigned to them. In a discussion on Carl Jung's concept of synchronicity, which is often wrongly taken to mean a simple simultaneity of events whereas Jung meant rather a strong psychological connection between events, Stephan Hoeller writes that:

Time and space are relativized in certain experiences of the psyche, which implies that their subjective reality is altered in accordance with meaning. The meaningful part of a synchronistic experience is never the mere connection of events in objective astronomical time, but rather in the power of meaning residing in the subjective image in the psyche, which

relates itself dynamically to an outer event.⁴⁶⁷

In my own life this has focused on certain ideas, and when I find myself coming back to them or being re-exposed to them (for whatever reason, or for no ostensible reason at all), I discover a correlative strong sense of having looped back to that time when the idea in question first made an impression on me. Only I have not ‘looped’ so much as ‘spiraled’, for in going back to the beginning I take with me my new perspective won through the experiences had in between. (I think that such might apply as well to instances of déjà vu.) These sensations of spiraling do still contain some aspects of a linear, physical sciences notion of time for I am mindful of being twenty-two and coming across Y or the like, but the felt dimensions are of another order entirely, and in that feeling there is not even a trace of a ‘gap’ between now and then – it is a poignant perception of direct connection stemming from the personal meaningfulness that has been assigned. In reflecting on the experience involved it strikes me as credible to speculate that these meanings are more emotional and intuitional in nature than they are rational, and hence that they are products of the System 1 layer of mind: the automatic, effortless, rapid, and preconscious level that underwrites the more laborious and inefficient awareness based reasoning of System 2. The meanings in question are nevertheless often able to be pinpointed in explicit thought using System 2, or at least so approached, except (I propose) in cases of déjà vu. There I think the coupling meaning is only given preconsciously in the mind and is accordingly inaccessible to a logical analysis, and hence the odd ‘there is something here but I cannot say what’ that accompanies such occurrences. Thus we phenomenologically feel our ways through life in more or fewer spirals, and though of course our bodies (including brains) age in a linear fashion and rot away as they do, our minds (emergent functioning) enjoy a

⁴⁶⁷ Stephan A. Hoeller, *The Gnostic Jung and the Seven Sermons to the Dead* (Wheaton, IL: Quest Books, 1982), p. 181.

different experiential trajectory.

That statement, and this position, is not meant in a dualist sense, however, nor is it drawing any dualist or even quasi-dualist conclusions; such would be ontic while the argument here is phenomenal. It might be helpful at this point if we remember to break out of the mind/body reflex which we have culturally developed and inherited by again turning to Bergson. He writes that experience tells us only that the mental and the physical are interdependent; from that an equivalency of the two does not follow⁴⁶⁸ (and, I might add, we should also take care about any further implications we may be tempted towards). We ought to recall as well just how historically situated our conceptual lives are. Expanding slightly on Bergson's discussion of mind/body in the Classical world related above, he tells us that dualism was then a non-issue precisely because the Idea informed all, making the Ancients' 'psyche' less spiritual than our 'mind' and their 'body' less corporeal than ours. The 'soul' or 'mind' was not defined by consciousness, nor the 'body' by extension: 'The scission was not yet irremediable between the two terms.'⁴⁶⁹ This is once more, I believe, an issue of the levels upon which we think and therefore speak (and therefore think again). I am not suggesting that there are 'dual times' which operate phenomenologically on us, instead I am attempting to argue that time in its singularity (whatever that might be ontologically) is experienced on these two planes: that of the physical sciences kind and that of the felt kind, of the absolute and the duration. Mind is a part of body – as we have been at pains to stress throughout this study –; it is a natural outgrowth of continual physical functioning just as a fingernail or a strand of hair is: all types of experience are contained in the same single organism, and all nuances of time are as well. The thinking mind-body in its operation knows time as spirals woven through with meaning, and the

⁴⁶⁸ Bergson 1998/1911, *op. cit.*

⁴⁶⁹ *ibid.*, pp. 349-350; quoted section from p. 350.

automatic mind-body in its operation knows time as linear and the corollary effects of decay, suffering, pain, and sometimes joy. The level of time that we focus on is a matter of purposive awareness, of the conceptual lying behind the expressed, and of the efforts that we engage in at moving from the received ‘made’ self to the ‘making’ self. That task, moreover, is one taken in the ever-now, one confirmed moment by moment.

6.3 Exhausted conclusion

This study has attempted to carry the argument that there is such a thing as the self, that it is a part of every human being (at least, but perhaps nonhuman animals as well; for reasons of scope and practicality we have focused only on the human situation), and that it is foundational to the higher particularizing levels of personal identity and whole person. The self informs these other levels but is also informed by them. We made the case that this self is composed of inherited defining elements such as those stemming from social, cultural, biological, epochal conditions and the like, and also from personally determined elements such as goals, aspirations, desires, dreams, hopes, beliefs, creeds, and similar features (labeled Set 1). Many of these items are either emotional in nature themselves or affect the emotions, and to them were added further emotions that are directly about the self via making comparisons (who I am versus who I want to be, how I am behaving versus the kind of behavior I think ‘a person like me’ should be doing, et cetera; labeled Set 2), and via an awareness about the self (taking the self as an attention focused and intentional object – what am I? what ought I to be involved in? et cetera; labeled Set 3). The foregoing together were our three ‘emotional’ sets, and to them we also added consciousness (in an underlying, always ongoing understanding of it; labeled C) and the presence of a physical body (necessarily so, but absent the particulars which instead go into personal identity; labeled B_p), for without those latter two no amount of any emotions would add up to anything, nor even be possible given natural demands. This yielded our self definitional

equation of: Set 1 + Sets 2 & 3 + C + B_p = the self. How did we arrive there? What basis for such a definition was given?

We firstly surveyed some of the very wide literature in the field of self studies, looking in particular at four representative types of the self: 1) the soft anti-realist position, 2) the hard anti-realist position, 3) the soft realist position, and 4) the contextualized soft realist position. The hard realist position (e.g. a Cartesian ego or a traditional soul form of the self) was not covered due to its already having been discussed in a great many other places and, as a result of those many analyses, largely disappeared from the academic world even if it remains in certain ways in some areas of thought or teaching. From our investigation of the four types we found that the soft realist position was the most initially robust of the grouping, and to it we also wished to add some of the elements stressed in the contextualized soft realist position. We therefore decided to take Kristján Kristjánsson's soft realist self-view, as presented in his *The Self and Its Emotions*,⁴⁷⁰ as our jumping off point into the deep waters of what a realist self might be and mean.

Finding the seas of the self choppy and the going rougher than we imagined, we decided to consult recent studies in psychological and neurological areas so that our self theory might be better informed about the background biological concerns affecting each of us as living creatures in a changing world. As part of this exploration we focused too on that aspect of world, of environment in its fullest sense, and discovered that the many embedded factors reaching into every nook and cranny of our lives extend a tremendous influence on the self as it forms and re-forms. It was at this point, after all of these considerations, that we proposed our definition of the self just listed.

⁴⁷⁰ Kristjánsson, *op. cit.*

Yet that was not found to be nearly enough. Having initially defined the self, and with it personal identity (self + contingent facts about the body (CF_B) + other contingent facts (CF_O) + feedback from the social realm (FB_S)) and whole person (personal identity (PI) + embeddness (E)), and from those definitions further established the two way relations between the three layers, we realized just how crucial one's emotions, intuition, and bodily factors are to the self, to identity, to experience, and to a great many areas of life as lived. Each of those items is naturally connected with phenomenological concerns, and so in order to further explicate them we next turned in their direction and continued our strenuous swim (though perhaps some might see this as only treading water). Within this examination, and as potential objections or counter theories to our account, we reviewed two other phenomenologically based self positions: that of Galen Strawson's thin self or minimal subject, and that of Barry Dainton's phenomenal self, both of which were reasoned to lack sufficient substance to ground the self as experienced based on the theoretical framework we had established by that point of our study. We recognized that we needed more than what either had to offer. To get that we returned to phenomenology's founder, Edmund Husserl, and refreshed our thinking on his ideas and important methodological practices before moving on to his famous student Martin Heidegger's work on the self. From these sources we were able to extend our picture while also adding beneficial depth.

Thanks in part to our journey through phenomenology we were awakened to the central place that consciousness plays for the self, and thus we shifted our stroke to a front crawl and carried on with our aquatic quest in the ocean of self. Consciousness necessarily being many leagues deep we strove hard against its undercurrents in a deliberation of panpsychism and that notion's connections with subjecthood and the self. Ultimately we found that the idea, as represented in Galen Strawson's thought (examined due to the

connection with his self theory earlier considered) was not tenable, but that what instead was defensible was the position that consciousness might always be present in a living human being. With that as aid we wisely cut back into the current proper and floated along into a review of cognitive science's data on consciousness. These were applied to the 'problem' of qualia while sustaining a focus on the self and its phenomenology, on what all that we were learning might mean for our own self-view. Finally, as we neared the far shore that we had set out for, we proposed a metaphysical basis for a version of reality that was ideational rather than empirical, but no less observable and applicable for that. Time, and time's reach into the self, was debated and an alternative conception of time's place for us, and purposefully lived by us, was offered. Stepping at last onto dry land and wiping the residual water from our eyes, we thought we had a soft realist self that might be worth turning over and, à la how Husserl would advise us to do, investigating its various sides, distinctions, qualities, and manifold details.

This resultant and hard won self was argued to be substantive and real, but real as an idea might be real, as a highly influential and directing concept might be real, and furthermore so informative that it actually becomes *forming*, with the outcome a psychological posit that we label as 'my self' and which goes on to continually address every aspect of our lives from the first hints of infant awareness to the final greying moments beheld by cloudy eyes. This self that is a self concept attains its degree of reality via its function, and we gave the notion of the 'functionally existent' in support. We see the self working and know it intimately but of course cannot assign a physical place or region to it – but such is not a rare condition in our world and mind/body discussions (amongst others) have provided many prominent examples of this phenomenon. There are few – perhaps no – ideas closer to the identities we so fret over in the shaping of our lives than that of the self, and so to pejoratively assign an epithet such as 'it is just an idea' to this concept would seem to have

entirely missed the point of our efforts.

This self does not exist in a vacuum housed within the whole person, and nor does the whole person herself exist in a vacuum devoid of others. The self, personal identity, the whole person, and everything else known and unknowable, is embedded in some kind of context, and the contours of those situational factors will ply their influences on the self far more than the self, through expression, will shape them. Yet the flow does run in both directions, and just as the world into which we tumble (or are thrown, in Heidegger's terminology) greatly affects us, so we affect it. Still, that over which we have no control exerts a very long reach into our lives, the consequence of which means that our choices are in many ways limited, particularly when we consider that for very much we will simply be unable to think it and therefore be blinded to it: our conceptual equipment gives rise to our perceived (potential) stimuli. As inhibiting as such factors are, however, there is a legitimate sense in which recognition enables release, in which simply learning of the way one's choices have been reduced greatly increases them. Although there is a staggeringly vast amount of aspects to our lives over which we cannot even fantasize about having any measure of authority or sway, our blessing and our curse is the awareness that we bring into this picture.

The problem – if it can be called that – is not only that we are alive, but that we are so aware of being alive, and of being ourselves as alive. It might be better for us if we had no self awareness at all – ignorance might be bliss, the debate is an extremely old one – but having that which we do forces us to face the question of how to live: as 'made' or as 'making' selves? This in turn involves our approach to time, with the standard position being the passive one wherein we comprehend time as purely external and unreflectively become carried along by our cultures, histories, circumstances (and now digital devices),

living more as manufactured and programmed machines than as dynamic agents. However the possibility of the 'making' self can begin, it was suggested, once more in time by shifting to the phenomenological ever-now, to the resolute and noticed now, to the time that can spiral through meaning and to the time where meanings allotted are done so purposefully and intentionally. If we ask 'How do I want to be?' and if we frequently follow that with 'How am I doing in that?' we are on the way, indeed deeply involved, in the never-ending composing of the self. We work at transformation, we work at becoming, and we work always *towards* that for there cannot be any final attainment of self completion nor any rest from self maintenance. The self is real but, like all else, it is ever changing as our emotions and the boundaries of our lives unfold in the shifting world around us. The 'making' self is an effortful carrying out, and as soon as we stop in our self building and self reflection the 'made' self invades and threatens to siliconize us into an unthinking automaton of finishing now this, now that; something many refer to simply as 'the daily routine'.

There is life and there is life, there is the 'made' self and the 'making' self. Yet the self is always there as the core of us, as that which we discover when we look, and as that out of which we act even when we do not. The self is the psychological foundation underwriting all, shaping and taking shape. How much of each will depend on our whole person choices, on our involvement and its direct or indirectness, on our awareness and its active or passive nature. There is of course no guarantee of success in any attempt at self direction, at self creation, but only in its pursuit do we find the chance of coming close to the potential inherent in the being of a human being. Surely that is a risk worth taking.

Bibliography

AFP-JIJI, “‘Smart’ tropical fish can recognize human faces’, The Japan Times: On Sunday, 26 June 2016, 21

AFP-JIJI, ‘Tasmanian devils pull back from brink of extinction’, The Japan Times, 24 September 2016.

<[Asma, Stephen T. and Greif, Thomas, ‘Affective Neuroscience and the Philosophy of Self’, *Journal of Consciousness Studies*, 19:3-4 \(2012\), 6-48](http://www.japantimes.co.jp/life/2016/09/24/environment/tasmanian-devils-pull-back-brbri-extinction/#>.WEpn8LlqMuc. Accessed 09 December 2016</p></div><div data-bbox=)

Backhaus, Peter, ‘The Japanese traffic light blues: Stop on red, go on what?’, The Japan Times, 25 February 2013.

<[Bakhurst, David, ‘Wittgenstein and “I”’, in *Wittgenstein: A Critical Reader*, ed. by Hans-Johann Glock \(Oxford: Blackwell Publishers Ltd., 2001\), pp. 224-245](http://www.japantimes.co.jp/life/2013/02/25/language/the-japanese-traffic-light-blues-stop-on-red-go-on-what/#.WUhlbISGOuk> . Accessed 20 September 2017</p></div><div data-bbox=)

Barresi, John and Martin, Raymond, ‘History as Prologue: Western Theories of the Self’ in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), 33-56

Bergson, Henri, *Creative Evolution*, trans. by Arthur Mitchell (Mineola, NY: Dover, 1998

unabridged republication/New York: Henry Holt and Co., 1911)

Bergson, Henri, *Time and Free Will: An Essay on the Immediate Data of Consciousness*, 3rd edn, trans. by F. L. Pogson (Mineola, NY: Dover, 2001 unabridged republication/London: George Allen and Co., Ltd., 1913)

Bermúdez, José Luis, 'Bodily Awareness and Self-Consciousness' in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), 157-179

Blasi, Augusto, 'Bridging moral cognition and moral action: A critical review of the literature', *Psychological Bulletin*, 88:1 (1980), 1-45

Caputo, John D., *Truth* (London: Penguin Books, 2013)

Carman, Taylor, *Heidegger's Analytic: Interpretation, Discourse, and Authenticity in Being and Time* (Cambridge: Cambridge University Press, 2003)

Cassam, Quassim, 'The Embodied Self' in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), 139-156

Chalmers, David J., *The Conscious Mind: In Search of a Fundamental Theory* (Oxford: Oxford University Press, 1996)

Coleman, Sam, 'Being Realistic: Why Physicalism May Entail Panexperientialism', in *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?*, ed. by

Anthony Freeman (Exeter, DEV: Imprint Academic, 2006), 40-52

Chudnoff, Elijah, *Intuition* (Oxford: Oxford University Press, 2013)

Dainton, Barry, *The Phenomenal Self* (Oxford: Oxford University Press, 2008)

Damásio, António, *Descartes' Error: Emotion, Reason, and the Human Brain* (New York: Putnam, 1994)

Damásio, António, *The Feeling of What Happens: Body and Emotion in the Making of Consciousness* (New York: Harcourt Brace, 1999)

Damásio, António, *Self Comes to Mind: Constructing the Conscious Brain* (New York: Vintage Books, 2012)

Dennett, Daniel C., *Consciousness Explained* (New York: Little, Brown and Co., 1991)

Descartes, René, *A Discourse on the Method*, trans. and intro. by Ian Maclean (Oxford: Oxford University Press, 2006)

Dijksterhuis, Ap, 'Think Different: The Merits of Unconscious Thought in Preference Development and Decision Making', *Journal of Personality and Social Psychology*, 87:5 (2004), 586-598

Eddington, Arthur, *The Nature of the Physical World* (New York: Macmillan, 1928)

Edlund, Matthew J., 'Sleep and Your Different States of Consciousness: Thinking may not be what you thought', *Psychology Today*, 13 January 2011.

<<https://www.psychologytoday.com/blog/the-power-rest/201101/sleep-and-your-different-states-consciousness>>. Accessed 23 November 2015

Escudero, Jesús Adrián, 'Heidegger on Selfhood', *American International Journal of Contemporary Research*, 4:2 (2014), 6-17

Feyerabend, Paul, *Against Method*, new edn, intro. by Ian Hacking (London: Verso, 2010)

Fischer, John Martin, *Deep Control: Essays on Free Will and Value* (Oxford: Oxford University Press, 2012)

Gallup, Gordon G., Jr, Anderson, James R., and Platek, Steven M., 'Self-Recognition' in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), 80-110

Gaus, Gerald F., *Political Concepts and Political Theories* (Oxford: Westview Press, 2000)

Gazzaniga, Michael S., *Who's In Charge?: Free Will and the Science of the Brain* (New York: Ecco Press, 2011)

Goff, Philip, 'Experiences Don't Sum', in *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?*, ed. by Anthony Freeman (Exeter, DEV: Imprint Academic, 2006), 53-61

Goldie, Peter, *The Emotions: A Philosophical Exploration* (Oxford: Oxford University Press, 2000)

Greene, Joshua, *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them* (New York: Penguin Press, 2013)

Gupta, Prahlad, and MacWhinney, Brian, 'Vocabulary acquisition and verbal short-term memory: Computational and neural bases', *Brain and Language*, 59 (1997), 267-333

Gusnard, D. A. and Raichle, M. E., 'Searching for a Baseline: Functional Imaging and the Resting Human Brain', *Nature Reviews Neuroscience*, 2 (2001), 685-694

Hacker, P.M.S., 'Is there anything it is like to be a bat?', *Philosophy*, 77:300 (2002), 157-174

Haidt, Jonathan, 'The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment', *Psychological Review*, 108:4 (2001), 814-834

Haidt, Jonathan, *The Righteous Mind: Why Good People Are Divided by Politics and Religion* (New York: Pantheon Books, 2012)

Hare, R. M., *Moral Thinking: Its Levels, Method, and Point* (Oxford: Oxford University Press, 1981)

Harman, Graham, *Tool-Being: Heidegger and the Metaphysics of Objects* (Peru, IL: Open Court, 2002)

Harrison, Scott, 'On the Analogy Between Mind/Brain and Software/Hardware', Massline.org, 04 December 1992; updated 24 June 1998. <<http://www.massline.org/Philosophy/ScottH/mindsoft.htm>>. Accessed 18 April 2017

Hatfield, Gary, *The Routledge Guidebook to Descartes' Meditations* (Abingdon, OX: Routledge, 2014)

Hawking, Stephen, *A Brief History of Time* (New York: Bantam Books, 1988)

Hegel, Georg Wilhelm Friedrich, *The Phenomenology of Spirit*, trans. by A.V. Miller, for. and comm. by J.N. Findlay, rev. edn (Oxford: Oxford University Press, 1977)

Heidegger, Martin, *Being and Time*, trans. by Joan Stambaugh, rev. and for. by Dennis J. Schmidt (Albany, NY: State University of New York Press, 2010)

Heidegger, Martin, *Basic Writings: Key Selections from Being and Time to The Task of Thinking*, ed. by David Farrell Krell and for. by Taylor Carman (New York: HarperCollins, 2008)

Henriksen, Birgit, 'Three dimensions of vocabulary development', *Studies in Second Language Acquisition*, 21 (1999), 303-317

Henry, Aaron and Thompson, Evan, 'Witnessing from Here: Self-Awareness from a Bodily versus Embodied Perspective' in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), 228-249

Hoeller, Stephan A., *The Gnostic Jung and the Seven Sermons to the Dead* (Wheaton, IL: Quest Books, 1982)

Hofstadter, Douglas, *I Am A Strange Loop* (New York: Basic Books, 2007)

Hume, David, *A Treatise of Human Nature*, intro. and ed. by David Fate Norton and Mary J. Norton (Oxford: Oxford University Press, 1739/2000)

Husserl, Edmund, *Logical Investigations, Volume 1*, trans. by J. N. Findlay, pref. by Michael Dummett, intro. and ed. by Dermot Moran (Abingdon, OX: Routledge, 2001)

Husserl, Edmund, *The Essential Husserl: Basic Writings in Transcendental Phenomenology*, intro. and ed. by Donn Welton (Bloomington, IN: Indiana University Press, 1999)

van Inwagen, Peter, *Material Beings* (Ithaca, NY: Cornell University Press, 1990)

Jackson, Frank, 'Epiphenomenal Qualia', *The Philosophical Quarterly*, 32:127 (1982), 127-136

Jackson, Frank, 'Galen Strawson on Panpsychism', in *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?*, ed. by Anthony Freeman (Exeter, DEV: Imprint Academic, 2006), 62-64

James, William, *The Principles of Psychology*, 2 vols. (New York: Dover Books,

1890/1950)

Kahneman, Daniel, 'A Perspective on Judgment and Choice: Mapping Bounded Rationality', *American Psychologist*, 58:9 (2003), 697-720

Kahneman, Daniel, *Thinking, Fast and Slow* (New York: Farrar, Straus and Giroux, 2011)

Klein, Gary A., 'A Recognition-Primed Decision (RPD) Model of Rapid Decision Making', in *Decision Making in Action: Models and Methods*, ed. by Gary Klein, Judith Orasanu and Roberta Calderwood (New York: Ablex Publishing Corp., 1993), 138-147

Klein, Gary A., *Sources of Power: How People Make Decisions* (Cambridge, MA: MIT Press, 1998)

Knowles, Lance, 'Mind blocks', *Language Magazine: Journal of Communication and Education*, 7 (2008), 28-34

Kriegel, Uriah, *The Varieties of Consciousness* (Oxford: Oxford University Press, 2015)

Kristjánsson, Kristján, *The Self and Its Emotions* (Cambridge: Cambridge University Press, 2010)

Kurzweil, Ray, *The Age of Spiritual Machines: When Computers Exceed Human Intelligence* (New York: Viking Press, 1999)

Lear, Jonathan, *Love and Its Place in Nature: A Philosophical Interpretation of Freudian*

Psychoanalysis (New Haven, CT: Yale University Press, 1998)

Legrand, Dorothée, 'Phenomenological Dimensions of Bodily Self-Consciousness' in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), 204-227

Locke, John, 'Of Identity and Diversity', in *An Essay Concerning Human Understanding*, 2nd edn (1689/1690), Chapter XXVII. Available on Project Gutenberg: <<http://www.gutenberg.org/cache/epub/10615/pg10615.html>>. Accessed 08 October 2015

Lombardi, Michael A., 'Why is a minute divided into 60 seconds, an hour into 60 minutes, yet there are only 24 hours in a day?', *Scientific American*. <<https://www.scientificamerican.com/article/experts-time-division-days-hours-minutes/>>. Accessed 24 November 2017

Lowe, E. J., *Subjects of Experience* (Cambridge: Cambridge University Press, 1996)

Luper, Steven (ed.), *The Cambridge Companion to Life and Death* (New York: Cambridge University Press, 2014)

Lycan, William G., *Consciousness and Experience* (Cambridge, MA: The MIT Press, 1996)

Manzotti, Riccardo and Parks, Tim, 'The Challenge of Consciousness', *The New York Review of Books: NYR Daily*, 21 November 2016. <<http://www.nybooks.com/daily/2016/11/21/challenge-of-defining-consciousness/>>.

Accessed 27 February 2017

Manzotti, Riccardo and Parks, Tim, 'The Color of Consciousness', *The New York Review of Books*: NYR Daily, 08 December 2016. <http://www.nybooks.com/daily/2016/12/08/color-of-consciousness/?utm_content=buffbu03030&utm_medium=social&utm_source=facebook.com&utm_campaign=buffer.>.

Accessed 27 February 2017

McGinn, Colin, 'Hard Questions: Comments on Galen Strawson ', in *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?*, ed. by Anthony Freeman (Exeter, DEV: Imprint Academic, 2006), 90-99

Mlodinow, Leonard, *The Drunkard's Walk: How Randomness Rules Our Lives* (New York: Vintage Books, 2009)

Musil, Robert, *The Man Without Qualities*, trans. by Sophie Wilkins and Burton Pike (New York: Alfred A. Knopf, 1995)

Nagasawa, Yujin, 'Thomas vs. Thomas: A New Approach to Nagel's Bat Argument', *Inquiry: An Interdisciplinary Journal of Philosophy*, 46:3 (2003), 377-394

Nagel, Thomas, 'Moral Luck', in *Mortal Questions* (New York: Cambridge University Press, 1979), 24-38

Nagel, Thomas, 'Panpsychism', in *Mortal Questions* (New York: Cambridge University Press, 1979), 181-195

Nagel, Thomas, 'What is it like to be a bat?', in *Mortal Questions* (New York: Cambridge University Press, 1979), 165-180

Nagel, Thomas, *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature Is Almost Certainly False* (New York: Oxford University Press, 2012)

Narveson, Jan, 'Libertarianism vs. Marxism: Reflections on G. A. Cohen's "Self-Ownership, Freedom and Equality"', *The Journal of Ethics*, 2:1 (1998), 1-26

Neisser, Ulric, 'Five Kinds of Self-Knowledge', *Philosophical Psychology*, 1 (1988), 35-59

Nelkin, Dana K., 'Moral Luck', *The Stanford Encyclopedia of Philosophy* (Winter 2013 Edition), ed. by Edward N. Zalta.

<<http://plato.stanford.edu/archives/win2013/entries/moral-luck/>>. Accessed 13 May 2014

Nietzsche, Friedrich, *The Portable Nietzsche*, ed., trans., and intro. by Walter Kaufmann (New York: Viking Penguin Inc., 1954)

Nietzsche, Friedrich, *Beyond Good and Evil: Prelude to a Philosophy of the Future*, trans. and comm. by Walter Kaufmann (New York: Random House Inc., 1966)

Nixon, Robin, 'Matters of the Brain: Why Men and Women Are So Different', *Live Science*, 01 May 2012.

<<http://www.livescience.com/20011-brain-cognition-gender-differences.html>>. Accessed 22 June 2015

Nozick, Robert, *Anarchy, State, and Utopia* (New York: Basic Books, 1974)

Oberg, Andrew, 'Thinking Unempirically', *Philosopher*, 03 July 2017.
<<http://philosopher.io/Thinking-Unempirically>>. Accessed 12 September 2017

Osman, Magda and Stavy, Ruth, 'Development of intuitive rules: Evaluating the application of the dual-system framework to understanding children's intuitive reasoning', *Psychonomic Bulletin & Review*, 13:6 (2006), 935-953

Panksepp, Jaak, 'The Basic Emotional Circuits of Mammalian Brains: Do Animals Have Affective Lives?', *Neuroscience and Biobehavioral Reviews*, 35:9 (2011), doi10.1016/j.neubiorev.2011.08.003

Papineau, David, 'Comments on Galen Strawson', in *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?*, ed. by Anthony Freeman (Exeter, DEV: Imprint Academic, 2006), 100-109

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

Peacocke, Christopher, *The Mirror of the World: Subjects, Consciousness, and Self-Consciousness* (Oxford: Oxford University Press, 2014)

Pippin, Robert B., *Hegel on Self-Consciousness: Desire and Death in the Phenomenology of Spirit* (Princeton, NJ: Princeton University Press, 2011)

Platek, Steven M., Keenan, Julian Paul, Gallup, Gordon G., Jr, and Mohamed, Feroze B., 'Where am I? The Neurological Correlates of Self and Other', *Cognitive Brain Research*, 19 (2004), 114-122

Ramachandran, V.S., *The Tell-Tale Brain: Unlocking the Mystery of Human Nature* (London: Windmill Books, 2011)

Ravven, Heidi M., *The Self Beyond Itself: An Alternative History of Ethics, the New Brain Sciences, and the Myth of Free Will* (New York: The New Press, 2013)

Rochat, Philippe, 'What is it Like to Be a Newborn?' in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), 57-79

Roholt, Tiger C., *Key Terms in Philosophy of Art* (London: Bloomsbury Academic, 2013)

Rorty, Richard, 'Ethics Without Principles', in *Philosophy and Social Hope* (London: Penguin Books, 1999), 72-90

Rovane, Carol, 'Branching Self-Consciousness', *The Philosophical Review*, 99:3 (1990), 355-395

Russell, Bertrand, *The Analysis of Matter* (London: Routledge, 1927/1992)

Ryle, Gilbert, *The Concept of Mind*, intro. by Daniel C. Dennett (London: Penguin Books, 1949, 2000)

Sadler-Smith, Eugene, and Shefy, Erella, 'The intuitive executive: Understanding and applying "gut feel" in decision-making', *Academy of Management Executive*, 18:4 (2004), 76-91

Seager, William, 'The "Intrinsic Nature" Argument for Panpsychism', in *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?*, ed. by Anthony Freeman (Exeter, DEV: Imprint Academic, 2006), 129-145

Sempera, Richard, *Rephrasing Heidegger: A Companion to Being and Time* (Ottawa: University of Ottawa Press, 2007)

Smith, David Woodruff, *Husserl*, 2nd edn (Abingdon, OX: Routledge, 2013)

Stapp, H.P., 'Commentary on Strawson's Target Article', in *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?*, ed. by Anthony Freeman (Exeter, DEV: Imprint Academic, 2006), 163-169

Stoljar, Daniel, 'Comments on Galen Strawson', in *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?*, ed. by Anthony Freeman (Exeter, DEV: Imprint Academic, 2006), 170-176

Strawson, Galen, 'Realistic Monism: Why Physicalism Entails Panpsychism', in *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?*, ed. by Anthony Freeman (Exeter, DEV: Imprint Academic, 2006), 3-31

Strawson, Galen, *Selves: An Essay in Revisionary Metaphysics* (Oxford: Oxford

University Press, 2009)

Strawson, Galen, 'The Minimal Subject' in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), 253-278

Strawson, Galen, 'Mind and Being – the Primacy of Panpsychism', in *Panpsychism: Contemporary Perspectives*, ed. by Godehard Brüntrup and Ludwig Jaskolla (Oxford: Oxford University Press, 2016), 75-112

Tononi, Giulio, 'Integrated information theory', *Scholarpedia: The peer-reviewed open-access encyclopedia* 10:1 (2015), 4164. doi: 10.4249/scholarpedia.4164. Accessed 12 December 2016

Tsakiris, Manos, 'The Sense of Body Ownership' in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), 180-203

Tversky, Amos, and Kahneman, Daniel, 'Judgment under Uncertainty: Heuristics and Biases', *Science*, 185:4157 (1974), 1124-1131

Vogeley, Kai, and Gallagher, Shaun, 'Self in the Brain' in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), 111-136

Williams, Bernard, 'Moral Luck', *Proceedings of the Aristotelian Society*, supplementary vol. L (1976), 115-135

Wilson, Timothy D., Dunn, Dana S., Kraft, Dolores, and Lisle, Douglas J., 'Introspection,

Attitude Change, and Attitude-Behavior Consistency: The Disruptive Effects of Explaining Why We Feel the Way We Do', in *Advances in Experimental Social Psychology*, ed. by Leonard Berkowitz (Orlando, FL: Academic Press 1989), 19: 123-205

Zahavi, Dan, 'Unity of Consciousness and the Problem of Self' in *The Oxford Handbook of the Self*, ed. by Shaun Gallagher (Oxford: Oxford University Press, 2011), 316-335

'Friedrich Nietzsche: Psychological illness and death (1889-1900)', Wikipedia: The Free Encyclopedia.

<https://en.wikipedia.org/wiki/Friedrich_Nietzsche#Psychological_illness_and_death_.281889.E2.80.931900.29>. Accessed 21 September 2017

'Henri Bergson', Wikipedia: The Free Encyclopedia.

<https://en.wikipedia.org/wiki/Henri_Bergson>. Accessed 29 August 2018

'Metric system', Wikipedia: The Free Encyclopedia.

<https://en.wikipedia.org/wiki/Metric_system>. Accessed 20 September 2017

'Moral Luck', Wikipedia: The Free Encyclopedia.

<http://en.wikipedia.org/wiki/Moral_luck>. Accessed 13 May 2014

'Researchers teach robots to "feel pain"', BBC News: Technology, 26 May 2016.

<<http://www.bbc.com/news/technology-36387563>>. Accessed 13 July 2016

'The X-Files', Wikipedia: The Free Encyclopedia.

<https://en.wikipedia.org/wiki/The_X-Files>. Accessed 14 December 2016

Pinterest.com. <<https://www.pinterest.com/pin/418342252857562657/>>. Accessed 02
February 2016