The Transcendence of Computational Intelligence

by

John Thornton

BBus (Commercial Computing), Griffith University, Australia (1994)

BSc (Hons), Griffith University, Australia (1996)

PhD (Artificial Intelligence), Griffith University, Australia (2000)

A confirmation document submitted in partial
fullfilment of the requirements of the degree of
Doctor of Philosophy

School of Humanities
Arts, Education and Law Group
Griffith University, Queensland
Australia

February, 2015

Abstract

The nature of the subject matter of this thesis makes it difficult to provide a concise summary of what I intend to write. My aim is rather to *discover* answers to the questions I am posing through the very process of writing. To begin therefore, I can only state the basic questions that are motivating my work: What is intelligence? What is computation? What is computational intelligence? What could it mean to transcend computational intelligence? Is such transcendence *demonstrably* possible?

Of course I have presentiments of answers to these questions. Firstly, in agreement with Roger Penrose, I would say that any procedure that can be realised as a Turing machine can be considered computational. The question for the thesis is whether we can abstractly consider the functioning of the human brain to be equivalent to that of a (very sophisticated) Turing machine. To answer this I shall be examining the latest hierarchical predictive coding models of the functioning of the human neocortex and asking to what extent such (computational) models can account for the capacities of human intelligence. In particular, I shall be enquiring into our capacity to have a direct knowledge of what it is to be conscious, and asking whether such knowledge provides evidence of our being able to transcend the purely computationally determined actions of our neural systems.

The confirmation document itself comprises an introductory chapter which provides an outline of the final thesis. At the moment I envisage a three-part manuscript. Part One is concerned with the problem of gaining *explicit* access to a state of pure, pre-reflective consciousness and thereby discovering what it means to have a *direct knowledge* of consciousness. In Part Two I plan to examine the computational foundations of intelligence in the functioning of the human neocortex, both from a neuroscientific and phenomenological perspective. Part Three is then intended to draw together Parts One and Two in a consideration of the possible transcendence of computational intelligence.

Following on from the introduction, Chapters 2-5 are all intended for inclusion in the final thesis. They cover Part One described above, with Chapter 2 providing a method of access, and Chapters 3-5 examining the work of Descartes, Schopenhauer and Husserl in relation to the material presented in Chapter 2.

Finally, the appendices present two conference papers and one workshop paper that I have written and presented during my candidature. These papers cover various aspects of the questions I intend to address in Parts Two and Three, including the relevance of Heidegger to contemporary cognitive neuroscience (Appendix A), whether the activity of the human brain is causally closed under laws that determine the local low-level functioning of neural populations (Appendix B) and how contemporary models of neocortical functioning can be mapped onto Husserl's phenomenological account of temporal consciousness (Appendix C).

Statement of Originality

This work has not previously been submitted for a degree or diploma to any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

John Thornton

February, 2015

Contents

Abstract i

Statement of Originality iii

Contents v

Acknowledgements viii

1 Introduction **1**

1.1 Investigating Consciousness 3

1.2 Thesis Outline 5

1.2.1 Access and Triangulation 5

1.2.2 Computational Intelligence 5

1.2.3 The Transcendence Computational Intelligence 9

1.3 Progress to Date 12

2 Gaining Access **14**

2.1 Two Moments of Consciousness 15

2.2 Stopping Thought 17

2.3 'Looking' and 'Seeing' 19

2.4 Direct Knowledge of Consciousness 22

3 Descartes **25**

3.1 The Method of Doubt 25

3.2 The Circularity of Descartes' Doubt 28

3.3 Descartes' Discovery 29

3.4 Perception and Reflection 32

3.5 The Problem of Conceptualisation 34

4 Schopenhauer **37**

4.1 Concepts and Percepts 39

4.2 The Concept of Will 43

4.3 The Identity of Embodiment 46

4.4 The Feeling of Willing 48

4.5 The Pure Subject of Knowledge 50

5 Husserl **57**

5.1 The Life-World and the Epochē of the Objective Sciences 57

5.2 The Phenomenological Reduction 59

5.3 The Reduction and Direct Knowledge of Consciousness 64

Appendices **67**

A An Essential Difference **68**

A.1 Reconstructing the Cognitive World 68

A.2 The Non-Reduction of Psychological Phenomena 70

A.3 The Cognitive Science Framework 72

A.4 Introducing Heidegger 73

A.5 Heidegger's Realism 76

A.6 The Conceptualisation of Human Agency 78

A.7 The Incomprehensibility of Physicalism 82

A.8 The Reversal 86

A.9 The Essential Difference 91

A.10 Heideggerian Cognitive Science 93

B The Phenomenological Negation of the Causal Closure of the Physical **95**

B.1 Preliminaries 95

B.2 The Paradox of Phenomenal Judgment 97

B.3 The Acquisition of Phenomenal Concepts 98

B.4 The Pre-Understanding of Pure Phenomenal Quality 101

B.5 Phenomenal Concepts and their Reference 103

B.6 Phenomenological 'Seeing' 106

B.7 The Validity of the Phenomenological Demonstration 111

B.8 The Phenomenological Consequences 113

C Hierarchical Temporal Intentionality **115**

C.1 Introduction 116

C.2 Intentionality and the Phenomenological Reduction 118

C.3 Phenomenology and the Neural Correlates of Consciousness 120

C.4 The Phenomenology of Internal Time Consciousness 123

References 127

Acknowledgements

The writing of this thesis has already threaded its way through three universities as I have sought clarity in what I wish to present and have also sought to gain an education to make up for not having studied philosophy at an undergraduate level. I first wish to thank Bruin Christensen and David Chalmers for persevering with me at ANU and attempting to bring some discipline and clarity to my philosophical expression. Next I would like to thank the philosophy staff at the University of Queensland who allowed me to attend and interrupt their classes, especially Phil Dowe, and particularly Deb Brown, who also took on the job of supervising me during 2012.

However, my deepest acknowledgment goes to my current supervisor, John Mandalios, who, as a friend and colleague, first introduced me to philosophy at Griffith University more than ten years ago. It has been his understanding of the true meaning of philosophy that has inspired me to continue with this project when I might otherwise have faltered . . .

Chapter 1

Introduction

The origins of this thesis go back to a brief meeting I had with Marvin Minsky at the 1996 Pacific Rim International Conference on Artificial Intelligence in Cairns. At that time I was embarking on a teaching and research career in artificial intelligence and was working on search algorithms for solving constraint satisfaction problems. I had little considered the broader philosophical import of my work and took the task of artificial intelligence to be entirely technical, i.e. the creation algorithms to solve problems that formerly could only be solved by means of human expertise and judgment.

Minsky was invited to the conference as a respected founding father of the artificial intelligence community and was giving a keynote speech outlining his vision for the future of the area. This speech included a trenchant criticism of all those philosophers of mind who considered human intelligence to be more than the functioning of a particularly complex computer algorithm.

At that time I possessed a largely implicit idea of what it means to be conscious and intelligent, grounded in several years of experimentation with various meditative practices. These experiments had already revealed to me the limited nature of what I then called 'normal waking consciousness.' In listening to Minsky, and speaking with him briefly after his talk, I saw that my direct experience of higher states of consciousness had no place in his view of human and machine intelligence.[1] Furthermore, when challenged, I found I had nothing definite to say against Minsky's position – his certainty, and the authority of objective science that lay behind it, left me divided. At one level I could see the rationality of explaining consciousness and intelligence in terms of the physical functioning of the brain. At another, I recognised that this rational/scientific

[1] This refers to Minsky's well-known theory that mind is formed out of the hierarchical combination and interaction of ultimately mindless, simple components (Minsky, 1988).

approach was on the wrong track, and yet couldn't explain exactly why.

In the intervening years, I have become increasingly absorbed in resolving these two levels of understanding into a coherent unity. The current thesis is an attempt to express the fruits of this long-standing meditation. However, I now understand that the path I have taken is not going to be easily understood or accepted by anyone schooled in contemporary Anglo-American philosophy of mind. I find myself in a difficult situation. In order to communicate, I must first set about dismantling the pre-understandings that would reject what I have to say out of hand. This is not something that can be done *within* the context of normal philosophical discourse. It is that context that is to be exceeded in a way that at present remains unclear. Even to give a general indication of what I have in mind is going to lead to difficulties. I find I have no appropriate language at my disposal. Whatever terms I may employ are already associated with the very philosophical pre-understandings that are to be put into question.

Nevertheless, a beginning must be made. My intention is to investigate the truth of the assertion that (true) intelligence is not (ultimately) computational. Here, the term '(true) intelligence' indicates I am not concerned with a theory of intelligence, but, in advance, I am asking whether there is intelligence that is independent of our human conceptualisations, and whether this intelligence is directly knowable or accessible to us in such a way that its presence can become self-evident. So I am not going to argue that there must be such true intelligence on the basis of indirect considerations (such as the observation of intelligent behaviour). Instead, I intend to investigate intelligence *directly.*

Of course, the question is *how* am I (are we) going to discover any kind of philosophical truth concerning the nature of (true) intelligence on the basis of a direct investigation? If there were such a method then surely it would be a matter of common knowledge. And yet the disorder of contemporary philosophy shows that the direct investigation and discovery of self-evident philosophical truth is *not* a matter of common knowledge. What we encounter are philosophical *theories,* constructions that elaborate the consequences of possible philosophical positions in an effort to decide which is the most *plausible.* Whereas we are concerned with finding out what it means to be intelligent on the basis of the direct evidence of immediate conscious experience.

To begin we must leave open the question of how consciousness and intelligence are related. This is something we intend to *discover* not to theorise over. What we are first seeking is a *direct* knowledge of *consciousness.* Knowledge of consciousness holds the possibility of being direct because consciousness is not an object in the usual sense of the word, in that I (the investigator)

am not separate from that which I am investigating.

The question of *how* I am to gain access to a direct knowledge of consciousness is main topic of the first part of the thesis.

1.1 Investigating Consciousness

Firstly, I shall begin by introducing the state of immediate or direct consciousness. This state can only be *negatively* indicated to ordinary reflective consciousness, because entering the state requires the *suspension* of such reflective consciousness. One way of achieving this suspension is to negate or withdraw from the normal identification with the active intentionality of thinking. Once I no longer take myself to be the thinker of the thoughts that pass through consciousness, I become free from the necessity of *having* to think. I can *refuse* thought. In such a refusal, thoughtful reflection on experience cannot arise – there is only pure experience, 'now.' That is the state of immediate or direct consciousness.

It is a central tenet of this thesis that in order to study consciousness one first must discover what consciousness *is.* To use the obvious analogy, to discover what colour *is* (as a pure quality) I do not think about it (like Jackson's Mary), I simply *look* at my visual experience.[2] Similarly, to discover what consciousness *is,* I must *look* at my conscious experience (in its totality). In both cases, the looking is not a thinking. And in neither case can I directly express what is 'seen' in objective terms. I can point to the red book, but I cannot point to its 'redness.' I simply 'know' what red is, as a pure quality. In seeing the red book, I have its redness present to me in an immediate intuitive fulfilment. To argue about this, once it is seen, is ridiculous.

However, consciousness is not like colour. It is not a sensory quality – it is not something that appears 'in' my experience. It is that by means of which experience is possible, including my experience of colour. So, in a quite literal sense, my knowledge of what it is to be conscious is a knowledge of no-thing, of no-quality. 'Thingness' is what consciousness reveals.

The question here, and the question that is explored throughout the first part of the thesis, is *how* the state of immediate or direct consciousness can be brought to language in such a way that there is an immediate knowledge of consciousness itself, and not of some reflected, theoretical simulacrum of consciousness. It is *because* consciousness is not an object, that we cannot

[2]This refers to Frank Jackson's famous thought experiment concerning Mary, the vision scientist, who lacked direct experience of colour vision (Jackson, 1982), and the question of the nature of Mary's knowledge of colour. (The underlying issue of the experience of phenomenal colour is dealt with in more detail in Appendix B.)

approach this question by means of an ordinary reflection. Ordinary reflection reflects on what consciousness *presents* and does so by means of a separation of itself into the subject who reflects and the object that is reflected. But consciousness is the ground of this separation of the subject and the object. As soon as I separate from a pure state of consciousness 'now' and attempt to reflect on it 'from the outside,' I immediately lose touch with it. I *use* consciousness to achieve the reflection – and in so doing I destroy the unity of my being conscious 'now' – I 'look' in the wrong direction, away from consciousness itself towards some objectified 'thing' that only *represents* my being conscious. My conscious presence is absorbed (lost) in the reflection and so cannot see itself in its act of reflection.

And yet, in being conscious 'now' *without* reflection, I am not without knowledge. I 'see' what I see, I am conscious of the room. I do not need to reflect. For example, I do not need to think that there is chair in front of me in order to have the experience of there being a chair in front of me. My thoughtful reflection does not *produce* the chair, it expresses what I was *already* conscious of. In that pre-reflective consciousness there is a direct knowledge of whatever it is that is being experienced. That is what it means to be conscious. If there were no such direct knowledge, I would be *unconscious.* The knowledge is direct because it is not reflected, i.e. it is not mediately known as a correspondence between some *partial* aspect of my experiential totality and some linguistically specified thought-token.

But the problem remains: my direct knowledge, in being direct, is for me alone. I may 'know' what it is to be conscious without reflection, but this knowing is mute, it *precedes* language (it is not a *private* language,[3] it is the experiential *foundation* of language). Only someone else, who already 'knows' this state, who can negate their reflective consciousness, and *recognise* what remains, can fully understand what is being referred to here. But the fact remains, such recognition is *possible.* Once the ground of direct or immediate consciousness is secured from reflection, then language can be used in a new way, to refer *back* to this state, and to be verified from *within* this state.

To make the situation clear, I am intending to use language in just this way – i.e. as a means to communicate with other co-philosophising individuals who themselves already understand and have access to a state of pre-reflective consciousness. Without such access, the central content of

[3]This refers to Wittgenstein's so-called private language argument (Wittgenstein, 2001), i.e. the contention that it is not possible for language to meaningfully refer to something that is entirely private or subjective. (The issue of how language can refer to pure phenomenal experience is investigated further in Sections B.1 and B.6 of Appendix B.)

the thesis will remain opaque, i.e. it will refer to states and experiences that the reader can only imagine but cannot ground in immediate intuitive fulfilment.[4]

1.2 Thesis Outline

1.2.1 Access and Triangulation

Given the importance (for understanding the thesis) of attaining to a state of pure thought-free consciousness, I shall be devoting the Chapters 2 to 5 to a more in-depth exploration of how this state can be accessed and explicitly recognised. Chapter 2 looks at how a normal state of reflective consciousness can be exceeded by means of attempting to *perceive* oneself thinking. I then examine the possibility of encountering a direct (unmediated, unreflected) *knowledge* of the resulting state of pure (thought-free) consciousness by expanding on the notions of 'looking' and 'seeing' introduced in Section 1.1.

Following Chapter 2, the plan is to demonstrate that this state of pure consciousness is not an invention presented here for the first time, but rather is a state that has been repeatedly discovered throughout the history of Western philosophy. To this end I shall consider (as examples) the investigations of *René Descartes, Arthur Schopenhauer* and *Edmund Husserl.* My claim is that each of these philosophers encountered the *same* state of pure consciousness which they then conceptualised and reasoned about according to their differing motivations and presuppositions. In revisiting and criticising these well-known conceptualisations my aim is to *triangulate* pure consciousness so that it *emerges* as what is essentially identical in each philosopher's vision. At the same time I will be exploring how such conceptualisation *ought* best to proceed, by considering how these previous accounts have gone *astray,* i.e. how, in different ways, each philosopher left the state of pure consciousness, and re-asserted the supremacy of their *reflective* consciousness.

1.2.2 Computational Intelligence

Having secured access to a state of pure consciousness, the second (and yet to be written) part of the thesis will investigate the relation between conscious experience and the functioning of the human neocortex. A central claim here is that disciplined *reflection* on the functioning of conscious intentionality (along the lines inaugurated by Husserl) provides *evidence* concerning the structure of the corresponding processes occurring in the brain.

[4]The phrase 'intuitive fulfilment' is borrowed from Husserl and will be clarified in Chapter 5.

To this end I shall be considering the recent development of hierarchical predictive coding models of neocortical functioning.[5] Such models see the neocortex as instantiating computational processes that search for and extract regularity from the streams of information flowing in and out of the cortical system. In detecting regularity, the neocortex develops an ability to *re*-cognise the structural regularities (forms) that it was responsible for extracting in the first place.

From the perspective of conscious experience, such forms are the events and objects we perceive, the concepts and thoughts we reflect on, and the actions we exhibit. From the perspective of physical theory, the process of form detection and emission is understood as a dynamical interplay between information streams moving up the neocortex and self-generated predictions moving down. These top down predictions determine how the neocortex expects its own states to evolve, and the interaction between these predictions and the bottom up inputs determine the *surprise* value of that input.

According to the currently influential *free energy principle* proposed by Karl Friston (2010), the primary function of the neocortex (and biological systems in general) is to minimise the surprise value (or *surprisal*) of interactions with the environment (i.e. with the source of the information streams) by learning to generate better predictions or by taking actions that fulfil existing predictions. Here the idea of a prediction is extended to include perception *and* action, such that an action *actively* predicts a future state of the environment by itself changing that environment.

The free energy principle provides an overarching framework for understanding the physical behaviour of the neocortex without proscribing exactly how surprisal can be minimised. Current theories have built on Rao and Ballard's proposal that message passing in the neocortex is primarily concerned with reducing prediction error by means of approximate probabilistic Bayesian reasoning occurring between hierarchically structured neocortical regions (Rao & Ballard, 1999; Lee & Mumford, 2003; Rao, 2004). The idea is that each region learns to generate (in real time) the input it receives from multiple lower regions and then feeds back this 'bigger picture' in the form of predictions of lower region activity. Higher and lower level regions then interact by means of error signals to arrive at the most probable answer to the question: what is the correct structure of the information I am processing now?

❧

In relation to the broader literature, this section of the thesis is best classified as a work of

[5] For a comprehensive review of the area see (Clark, 2013).

neurophenomenology, i.e. it is an investigation of the fundamental relationship between phenomenology and cognitive neuroscience. As such, it revisits several of the questions raised by Rick Grush in a series of papers from the mid-2000s (Grush, 2004, 2005, 2006). Of particular relevance is Grush's criticism of the work of three pioneers of neurophenomenology (Francesco Varela (1999), Timothy van Gelder (1996) and Dan Lloyd (2002, 2004)) for their lack of precision in determining how temporal consciousness could be instantiated neural tissue. Grush's own Kalman filter emulation model provided an example of the kind of detailed computational process that the brain would need to instantiate (at least to underpin Husserl's tripartite temporal model (1991)). More recent work on hierarchical predictive coding has broadened Grush's idea of a brain-based emulation module to propose that the *primary function* of the neocortex is to minimise prediction error in the sensory input and motor output streams (as described above).

However, I shall be arguing that current predictive coding models still fail to provide an adequate (physical) explanation of the universal structure of temporal experience. Such experience is characterised by immediately present *temporal horizons* that hold on to past experience (via *retention*) and form corresponding expectations of future experience (via *protention*). In order to physically ground temporal experience I shall introduce Jeff Hawkins' *hierarchical temporal memory* (HTM) model of neocortical function (Hawkins & Blakeslee, 2004; Hawkins, George, & Niemasik, 2009; Hawkins, Ahmad, & Dubinsky, 2011) on which I have been working with my research group (the Cognitive Computing Unit) since 2006 (Thornton, Gustafsson, Blumenstein, & Hine, 2006; Thornton, Faichney, Blumenstein, & Hine, 2008; Thornton, Main, & Srbic, 2012; Thornton & Srbic, 2013). Hawkins' model differs from hierarchical predictive coding by implementing lateral temporal sequence learning within all regions of the neocortex (rather than relying only on top-down feedback to provide temporal context). This makes the brain's representation of the world entirely temporal (i.e. it deals only with sequences) and illuminates how and why the neocortex should (in each moment) preserve information about its previous activity, while also predicting its future activity. All this occurs within a recursive hierarchical structure that mirrors Husserl's notion of horizons of retention and protention surrounding the flowing primal impression of 'now' (Husserl, 1991).[6]

❧

[6]That is not to imply that we fully accept Husserl's tripartite model of temporal consciousness. As will become clear in Chapter 5, Husserl's notion of a primal impression is a problematic construction based on a reflected experience rather than on a direct knowledge of consciousness 'now.'

The *philosophical* import of this detour into brain science is to provide a clear and physically grounded notion of what is meant by *computational intelligence,* i.e. that form of intelligence whose functioning is determined by the physical processes occurring in the human brain.[7] I term this *computational* intelligence because I take it that the *normal* processes that occur in the human brain are determined by laws that (in principle) could be implemented as a Turing machine (i.e. as a computer program).

My assumption is that the internal (computational) functioning of the neocortex is mirrored in the underlying structures and processes we can directly observe in conscious experience. Such observation requires the taking up of a phenomenological stance in relation to immediate experience such as Husserl formalised in his account of the phenomenological reduction (and such as we have recognised as a state of pure consciousness). What modern neuroscience *adds* to Husserl's detailed phenomenological analyses is an alternative view of the genesis of the intentional structures that Husserl identified. This view brings into question Husserl's conceptualisation of intentionality by providing an entirely physical account of the syntheses involved in constituting a unified temporal experience of the world. More specifically, given that we can now trace the acts of synthesis that Husserl attributed to the intentionality of a transcendental subjectivity *directly* back to the physical activity of populations of neurons, can we still retain Husserl's conception of a transcendental subjectivity that (actively) *constitutes* experience? Do we, for example, picture a transcendental subjectivity whose activity *is* the activity of neural populations? In which case, how is this activity determined? Is the transcendental constitution itself determined by the lawful (computational) interactions of (unconscious) neurons?

At the same time, we should remember that from the perspective of the state of pure consciousness, neurophenomenology, as it is understood within the domain of contemporary cognitive science, is only indirectly or reflectively concerned with consciousness itself. In Husserl's language, it is a form of intentionalistic *psychology* that retains the naturalistic world view of the positive sciences by identifying consciousness with the immanent experience of particular physical bodies living within an independently existing spatio-temporal universe.[8]

Such naturalistic neurophenomenology has its place in the development of theories concerning the functioning of the human brain, but it does not encounter the transcendental dimension of

[7] Here the use of the terms 'physically grounded' and 'physical processes' does not imply the acceptance of a physicalist ontology, it refers to the third-person objectified *view* of the world that forms the basis of scientific investigation.

[8] For an overview of the project of naturalising phenomenology see (Petitot, Varela, Pachoud, & Roy, 1999).

consciousness. The underlying issue we are investigating in this thesis is the nature of the relationship between consciousness and the human brain, i.e. whether consciousness (intelligence) transcends the functioning of the brain, and whether I can have a direct knowledge of this transcendence. In order to investigate the possibility of transcendence I must (with Husserl) put all scientific/naturalistic assumptions out of play.[9] I simply do not know how my experience of there being an objective world has come into being or in what sense that objective world 'exists' independently of my consciousness. Naturalism must be put out of play because it *presupposes* an answer to a question it is unable to seriously investigate.

Nevertheless our discovery of the state of pure consciousness does not entitle us to *ignore* the findings of neuroscience when it comes to conceptualising what is discovered. A knowledge of what is occurring in the brain should at least be able to *assist* in such conceptualisation (for example, in not ascribing consciousness with transcendental powers that can be explained purely on the basis of lawful physical causation). The task here is rather one of finding a *balance* – one that recognises and synthesises both a direct knowledge of consciousness and our objective, empirical knowledge of how the brain functions.

1.2.3 The Transcendence Computational Intelligence

As should now be clear from these preliminary considerations, the thesis is primarily a work of investigation and discovery. The current introduction only outlines a direction for this investigation and a basic method of discovery. That means our destination remains open and, for now, I can only indicate where the line of enquiry of the second part the thesis is tending.

Firstly, there is the question of the meaning of Husserl's phenomenological investigations. From the state of pure consciousness I find experience is unified. Only in reflection, in making an object of experience, can I begin to distinguish the structures that Husserl has conceptualised. If I take experience exactly as it presents itself, it presents the world (the one world for both you and I) and my being in that world (and your being in that world).[10] I do not find a purely immanent layer of experience. I find I am already in contact with a world that lies outside me. To understand my experience as something immanent, I have to *deliberately* take up an *attitude* and then understand that experience *as* immanent.[11] This is *not* what occurs in my withdrawal of identification from

[9] This refers to Husserl's phenomenological epochē (Husserl, 1970/1992, Section 35).

[10] Here I am also thinking of Heidegger and his characterisation of being-in-the-world as the true starting point for phenomenology (Heidegger, 1962/2008).

[11] The point being that Husserl does analyse experience in terms of there being *immanent* conscious contents that

thought.

Nevertheless, we can still discern in Husserl's phenomenological reduction evidence that he too has encountered a state of pure consciousness.[12] This shows up in his attainment of a transcendental state of pure witnessing that is no longer 'human.'[13] It is from this state Husserl 'sees' that in our normal state of consciousness (the 'natural attitude') we experience (*believe*) ourselves to be 'men,' i.e. particular human individuals with particular private psyches whose contents are determined by our physical/sensory interactions with an external reality. In 'escaping' from the natural attitude, Husserl 'sees' it for the first time. In our language, he encounters a state of pure consciousness, and in so doing he encounters a direct knowledge of the *impersonal nature* of pure consciousness. He dis-identifies with his human self and discovers that the entire drama of being a human individual experiencing an external reality is encompassed by a pure consciousness that does not appear within the experience, but rather *presents* it, in the sense that it *gives meaning.* Here we cannot say that consciousness gives meaning *to* experience, as if experience were somehow already there but devoid of meaning. It is this very *giving* that *creates* experience. And yet, at the same time, it is possible to 'know' that I *am* that consciousness (cf. Husserl's "disinterested" spectator (Husserl, 1970/1992, p. 157)).

The kernel of Husserl's reduction is therefore not an act of *abstract reflection,* it is an act of disidentification. Abstraction is an *artificial* separation from immediate experience, whereas disidentification (from one's human individuality) brings about an immediate unification of consciousness with itself. Such disidentification is not an artificial taking up of a position in relation to experience, it is the *discovery* that an identification is already occurring. Once it is discovered, the identification can no longer operate as it did before – the discovery *negates* the identification and so reveals it *as* an identification. This revealing *is* a direct knowledge of consciousness.

However, from the perspective of the second part of the thesis, i.e. from our consideration of the relationship between Husserlian phenomenology and contemporary brain science, it turns out that we could equally well have pursued our enquiries from *within* a normal state of consciousness. The discoveries of the basic structures of temporal experience (of retention and protention), of the rules whereby the harmonious unfolding of temporal experience comes to be accepted as a perception of a material object, and so on, do not require that I disidentify with my human in-

are synthesised by means of an intentionality that constitutes *transcendent* intentional objects.

[12] This will be covered in detail in Chapter 5.

[13] It is Eugen Fink, in the Sixth Cartesian Meditation, who particularly emphasises that the phenomenological reduction involves the transcendence of human individuality (Fink, 1995).

dividuality. I simply need to abstract myself from experience and treat it as something flowingly immanent to consciousness. I can still understand this immanence as something produced by the activity of a human brain, as something that belongs to my physical human individuality. No realisation of a state of pure impersonal consciousness is required.

From this naturalistic perspective, the basic structures of intentionality that Husserl uncovered by means of his phenomenological investigations are simply phenomenal *reflections* of certain processes occurring in the brain. These processes are not intentional in themselves, but they are *experienced* as intentional.

So how are we to understand this reflection of intentionality? To begin, I would say that the brain processes that underlie my experience of intending a particular perceptual object are straightforwardly computational. That means I *involuntarily* intend what is intended according to rules that determine the (local low-level) behaviour of my neurons. And yet my being conscious *bestows meaning* on this neural activity. So the (computational) activity determines what I *must* intend, but it is my intending that *creates* the experience (the meaning) that there is something 'out there' in 'external space' in front of 'me' 'now.'

The deeper question, the one that lies at the centre of the thesis, is asking whether *all* my experience can be accounted for in this way, i.e. as the involuntary interplay between a computational system (the brain) and an impersonal meaning bestowing consciousness, whereby the meaning bestowing acts of consciousness are completely determined by the behaviour of the computational system. If such be the case then I would conclude that my *intelligence* is computational through and through and that all my insights and ideas are the straightforward consequence of the operation of what we now call the laws of physics. In contrast to this, I am going to claim that, not only does my intelligence transcend the low-level computational constraints of neuroscience, but that it is possible to realise (i.e. to have a direct knowledge) that this is the case.

The main task of this final section of the thesis is therefore to demonstrate the immediate and indubitable non-computational nature of pure intelligence. Perhaps the one Western philosopher who has come closest to addressing this question already is *Martin Heidegger.* I am reminded of his long meditation on 'What is called Thinking?' his asking 'what calls for thinking' and his insistence that the most thought-provoking thing is that we are *not thinking yet* (Heidegger, 1968/2004). From the perspective of the current work, Heidegger was seeking a way beyond the computational assembly and manipulation of concepts that characterises what we have now come to accept as philosophical thinking. His talk of 'getting under way' is for us a stepping out of

the computational frame of ordinary thought, a reaching towards a transcendental insight. Such insight is a creation of a meaning that cannot be reached by any rule-following process. Or, at least, this is what I intend to demonstrate.

1.3 Progress to Date

I have currently completed three further chapters, all aimed at providing a direct knowledge of what is meant by the state of pure consciousness. The first of these (*Gaining Access*) gives a practical demonstration of how to enter a state of pure consciousness by means of negating thought and then explains how we may gain direct (pre-reflective) knowledge of this state by means of a questioning that 'looks' and 'sees.'

In the next chapter (*Descartes*), I demonstrate how Descartes' method of doubt leads to the same state of pure consciousness already accessed by means of thought negation. I then consider how Descartes began to conceptualise this discovery on the basis of his pre-existing understanding of experience and so covered over his original insight.

In Chapter 4 (*Schopenhauer*) I examine the work of Arthur Schopenhauer, both as an example of post-Kantian philosophy, and because he clearly recognised the key (phenomenological) distinction between perception and conception that Descartes missed. Schopenhauer's realisation of this distinction enabled him to look more deeply into the state of pure consciousness but did not free him from his basic identification with his human self (i.e. his personal will).

The idea of these first chapters on Descartes and Schopenhauer is to bring the state of pure consciousness into relief by showing how it can become obscured once we start to *think* or *reflect* on the meaning of the state. It is in the next (unfinished) chapter on Husserl that we first encounter a philosopher who fully appreciates what has been discovered and who recognises that the state of pure consciousness cannot be treated as an ordinary object of thought – it rather must be approached and investigated on the basis of a disciplined method that understands itself and the state it is intended to encounter. This method (the phenomenological reduction) therefore becomes the object of a detailed investigation, both in terms of its similarities and its differences to our own approach of thought negation.

Once the chapter on Husserl is complete this will mark the end of Part One of the thesis. Although work has yet to start on the chapters for Parts Two and Three I have been examining and writing about the issues I intend to cover for a number of years. This has resulted in

the presentation of one workshop paper and two conference papers during the progress of my candidature. The first of these was written in conjunction with Bruin Christiansen from ANU and presented at 'Reconstructing the Cognitive World: A workshop with Michael Wheeler' at the Goethe University Frankfurt am Main in 2010. This paper is relevant to the thesis as it examines the significance of the philosophy of Martin Heidegger to contemporary cognitive science. In particular, the paper criticises Michael Wheeler's attempt to naturalise Heidegger and make him a spokesperson for his critique of 'Cartesian' cognitive science. I include the original version of this paper (solely authored by myself) in Appendix A. The final paper presented at the workshop was co-authored with Bruin Christiansen and can be obtained from: http://www.ict.griffith.edu.au/~johnt/publications/Wheeler2010.pdf

The second paper (see Appendix B) was presented at the 2012 conference of the Australasian Association for Philosophy at the University of Wollongong, and addresses the central issue of thesis by arguing that our ability to recognise and conceptualise pure phenomenal quality demonstrates that our intelligence is not (entirely) determined by physical law.

Finally, in Appendix C, I begin an explanation of how Jeff Hawkins' hierarchical temporal memory model can be mapped onto a Husserlian tripartite model of temporal consciousness by considering the sequence learning activity of minicolumns in the neocortex. This material was presented at the 18^{th} conference of the Association for the Scientific Study of Consciousness at the University of Queensland in July 2014. The material in the appendix covers the first half of the talk and is supplemented by the talk overheads which can be downloaded from: http://www.ict.griffith.edu.au/~johnt/publications/ASSC18Talk.pdf

Chapter 2

Gaining Access

Ordinary knowledge (common knowledge), at a first approximation, and in a sense that will be clarified as we continue, is only an *indirect* knowledge of consciousness. This includes all knowledge gained from introspectively reflecting on consciousness. Here the indirectness arises from the reflective intention to make an object of an experience. Such a procedure *divides* experience: there is the introspecting attentive process and that which it thematises in its introspection. The *total* experience is one of introspecting on an *aspect* of experience. I find I cannot reflect on the total introspective experience because introspection itself is *turned away from itself.* It can only thematise another concurrent experiential process, from which it has separated itself. That is not to imply that reflective introspection is somehow invalid. What we are asking is whether it is possible to have a direct knowledge of consciousness as a totality or unity – in a way that does *not* make consciousness an object of reflective introspection.

So what is it to be conscious? This is not a *theoretical* question. I mean to bypass the professional philosophical conceptualisations, the kind that would want to know, for instance, what kind of consciousness I am talking about, and would then expect me to clarify the concept I have in mind. I do not yet have a precise concept in mind. I am asking about this very consciousness now, and now, and now. And I am suggesting that we already *implicitly* know what it is to be conscious, on the basis of *being* conscious. It is not that knowledge of consciousness is somehow at a great distance and the discovery of this knowledge is going to require stringent efforts in terms of studying something unfamiliar. On the contrary, the difficulty is that my being conscious is *too* familiar – it is not something I can separate from myself – wherever *I* am, as a reflecting, perceiving intelligence, consciousness is, as well.

So the issue is not how to encounter being conscious – my being conscious already takes

care of that. The issue is how to transform my *implicit* knowledge of consciousness into *explicit* knowledge, or rather, how to *realise* this knowledge in such a way that it can be a self-evident truth rather than something I believe (or disbelieve) or hold to be probable. Again, we must recall that this is *not* a matter of *reflecting* on my implicit knowledge. Our task is to encounter being conscious, directly, that is, consciously, but *without* reflection – because in reflection we turn away from the actual experience, now. It is this turning away that is of interest.

2.1 Two Moments of Consciousness

To begin, I shall simply state that my being conscious is always occurring 'now.' If I consider being conscious in the past then I am thinking about an experience that was once conscious but is *not* conscious 'now.' Thinking about the past is an *act* of remembering, of being 'conscious of' a past experience, where the act itself is (implicitly) conscious 'now.' Even that of which I am conscious – my particular memory – is actually present to me 'now.' It is only the meaning-reference of the memory that connects with the past, that makes the memory an experience *of* the past.[1]

Each act of consciousness is 'conscious of' something and each act has its own particular kind of 'something' of which it is conscious. For example, an act of remembering is conscious of a memory, an act of perceiving is conscious of a perception, and an act of reflection is conscious of a reflected experience. The *first* moment of conscious experience comprises this being conscious of an experiential 'something.'[2] However, the act itself (e.g. the perceiving, the remembering) is always occurring 'now.' Encompassing this entire structure is an implicit consciousness that does not proceed from any act, but rather unifies and illuminates all my acts and that of which they are conscious. This implicit, synthesising, impersonal, illuminating consciousness is the source of there being any 'now' within which I can be conscious of anything. It does not appear *in* 'now' (like a consciousness of . . .), it *is* 'now.'

In distinguishing this implicit encompassing impersonal consciousness from our personal act-

[1] At this point, we cannot fully consider what occurs in an act of remembering, because we have yet to introduce the phenomenological concept of *intentionality*.

[2] I should emphasise that this is a high level and simplified characterisation. In normal experience the 'something' of which I am conscious is hardly ever a single memory or perception, but a much more complex phenomenon, with elements of perception, memory, and horizons of expectations all mixed together. We shall examine this in more detail when we encounter Husserl's account of internal time consciousness (Husserl, 1991).

based consciousness of the world, we now have *two* moments of consciousness. These moments comprise our everyday experience of being awake and directed towards the objects of the world. In being so directed, I am directed *away* from the encompassing consciousness 'now.' This remains true even for ordinary acts of perception directed towards the immediate 'now' of the world. For example, in being conscious of the chair in front of me 'now,' I am still only conscious of an object appearing *within* an encompassing consciousness, while the encompassing consciousness itself remains unattended.

What has been explicitly stated above concerning the two moments of consciousness is *already* coming from a direct knowledge of consciousness. It is only on the basis of having (somehow) directly encountered an encompassing consciousness 'now' that the two moments of consciousness can be discerned. The idea is that such explicit statements, insofar as they are true, will cause the corresponding implicit knowledge to respond or awaken. Whether such an awakening actually occurs depends (in part) on whether the explicit knowledge contradicts some currently held belief. If so, then that belief first has to be consciously examined. The intention is not that statements of explicit knowledge are taken up as beliefs. The idea is to *gain access* to a direct knowledge of consciousness. Such direct knowledge, in its very self-evidence, makes belief redundant (this will become clearer in our subsequent investigation of belief).

The distinction we are making is between two moments of consciousness that combine to form the inter-related unity of everyday consciousness. This inter-relationship is such that one moment (my explicit consciousness of ...) obscures the other (the encompassing consciousness now). What is needed, in order to directly encounter the encompassing consciousness, is a withdrawal from the normal stream of conscious experience that remains turned away from this encompassing consciousness (i.e. because it is directed towards the world). Such a withdrawal cannot be achieved by a simple redirection of my attention so that I am directed towards consciousness now – that is the path of reflective introspection, which, as we have already shown, cannot reach to a direct knowledge because it remains divided from the object on which it reflects. What is needed is a withdrawal from the very activity of being directed towards anything. But how is such a withdrawal to be achieved? Am I to somehow direct myself to not direct myself towards anything?

2.2 Stopping Thought

The solution to this paradox requires that we find a way that goes beyond our normal everyday state of consciousness. How this can be achieved is a problem usually taken to lie outside the domain of philosophy. Nevertheless, one of the central claims of this thesis is that no (true) philosophy is going to be possible from within the confines of a normal everyday state of consciousness, because such a consciousness remains excluded from the very domain it seeks to investigate. Furthermore, I am taking it that philosophical contemplation – true philosophising – is always, and has always been a matter of first escaping from the confines of the prevailing everyday state of consciousness. The difference here is that we are facing this task head-on, so to speak, rather than indirectly reading about the discoveries and insights of someone else.

To restate the situation, I am attempting to bring the implicit consciousness that accompanies my explicit consciousness of the world to a direct realisation. It appears I cannot do this (directly) by an act of will. So I must take an another route. The way I am proposing is to consciously observe myself while I am engaging in an act of reflection, i.e. while I am *thinking.* To think, as far as this experiment is concerned, is to be directed towards an object of thought, where an object of thought is understood in contradistinction to an object of *perception.* The crucial feature of an object of perception is that it is experienced as being actually (bodily) present 'now' i.e. in the 'now' of my encompassing consciousness. Hence we say that when we perceive something, we are immediately or directly conscious of it. In contrast, when I think of something, I am directed towards an object of thought which either refers me away from 'now' (e.g. towards something I remember or imagine) or towards an object of perception (i.e. something I perceive in front of me now). In both cases, thought is only indirectly related to 'now,' via the medium of past, present or future possible perceptions.

However, it certainly appears to the one who is thinking that I can think about *now.* For example, I can direct my thinking intention towards an object in front of me and I can think of the object, as it is, in front of me, 'now.' In that case the object of my thought, the meaning-intention, or reference of the thought, is the very object in front of me 'now.' However, when I actually direct myself in this way, I find I am no longer *thinking* of the object, I am *perceiving* the object. Any thought I may have of the object directs my attention away from the immediate perception onto a thought-token (a concept) that merely *represents* the perceived object. Of course, I can at any moment 'cash in' the representation in a direct perception of the object, but then I am no

longer thinking of the object.

So, if I attempt to entertain an object of thought that refers directly towards 'now' I will no longer be thinking, I will be perceiving. Similarly, if I attempt to reflect on an object I am perceiving 'now,' I will no longer be perceiving, I will be thinking. This is not an artificial distinction. To think and to perceive are recognisably different acts whose difference can be demonstrated in immediate experience. The test is to attempt to perceive oneself thinking, i.e. to consciously observe one's own act of thinking about something (anything) in the moment 'now' that one is actually thinking (just as one perceives a worldly process occurring 'now,' such as my raising my finger 'now'). To be clear, the task is not to observe what one was thinking 'just now.' That is an observation of something that has already occurred, and so is an act of thought directed away from 'now.' You have to catch (perceive, become conscious of) yourself thinking 'now.'

This task is analogous to that of a Zen koan, i.e. something inherently paradoxical and inaccessible for an ordinary state of consciousness, and yet something, once entered into with seriousness and sincerity, capable of moving one beyond that very state. If it works, the attempt to consciously observe yourself thinking will cause you to stop thinking. It is like two mirrors facing each other. So long as the mirror is reflecting on 'something' then it is conscious of that something, but if it attempts to reflect on itself reflecting, it will only find an empty mirror. There will be nothing, at least no thought, and in there being no thought there will be no one there to reflect on the fact that there is no thought. There will be a gap. In that gap, if the experiment succeeds, there will be consciousness without thought, without reflection.

Immediately *after* any gap you will probably think 'I wasn't thinking just then.' In having such a thought you return to a consciousness of the previous moment and leave the immediate consciousness 'now.' You may even think you imagined you were conscious without thought, because it is by means of thought that we make things memorable to ourselves. In order to verify that you were not thinking it is not enough to visit once and then remember the visit. Every memory is an indirect consciousness of ... The task is to *realise* that no thinking is occurring, that there is a pure thought-free, non-reflective state of consciousness, in the actual moment of such a state of consciousness. Such a realisation (clearly) cannot be a matter of thought, because, in thinking, one has left the thought-free state. Rather, one must *look.*

2.3 'Looking' and 'Seeing'

In order to investigate a thought-free state, I first require the ability to *voluntarily* shift my centre of consciousness away from the normal reflective (thoughtful) state of world understanding to a state of pure (thought-free) experience. To disengage in this way requires a *willingness* to disengage. If one is simply curious, seeking to obtain information to 'turn over' in one's mind, then no such disengagement is going to take place. Stopping thinking involves the gathering together of one's entire resource of consciousness so that it becomes immersed in what is present 'now.'

To begin, what is present 'now' is what I (passively) perceive 'now.' For example, there are the objects and movements appearing in my visual field, the sensations occurring in my body, the sound of the traffic, the birds, the wind in the trees, and so on. These events are *markers of the present,* of 'now,' and can be used to pull the consciousness back each time it gets caught up in a train of thinking. For if I remain absolutely present with what is occurring 'now,' I *cannot* think. It is only in being distracted from 'now' that thinking can get underway.

With practice, it becomes possible to voluntarily 'turn off' thinking by one-pointedly attending to whatever is happening 'now.' However, our aim is not just to enter into a thought-free state; we are seeking a *direct knowledge* of this state, i.e. a knowledge that is not the product of reflecting on a past 'now' but one that 'looks' directly into 'now,' *now.* In order to achieve this, I must still remain absolutely present to the immediacy of the world of my senses. It is this world that we now wish to investigate, to 'look into.' To get started, I shall simply state what I 'see' when I 'look' into this world, and leave the question of how such looking and seeing is achieved until we have a grasp on the domain that is being seen.

The situation 'now' is that I have suspended all thinking and I am simply present in the world of my senses. In not thinking, I straightforwardly accept this sensory world to be exactly what it presents itself to be. I leave behind all my abstract interpretations, such as thinking that what I 'really' experience is something constructed in my brain. I become naïve. The sun rises. The chair stands before me and it is blue. I have no notion of photons impinging on my retina, or of my experience being composed of sensations. I do not frame what I experience as being immanent or subjective. I have not yet split my world into subjective and objective components. I do not even frame my experience as being an 'experience.' I am simply 'here' in this room and the things in the room are 'here' in front of me. What I see are *the things themselves, just as they appear.* I implicitly understand myself as being here, present in and to the world, already and directly

immersed in a shared reality. To think of this understanding as something subjective is to have already separated from the immediacy of 'now' and to have started to *reflect.* Such reflection overlays my implicit and immediate thought-free state. For instance, *before* I reflect, I 'know' that the chair, the actual chair, is present before me, that what I see *is* the chair - my 'knowing' this is already embedded in my unreflected consciousness of the chair. As soon as I reflect, my entire frame of reference *shifts.* Now I can think of my seeing the chair as something subjective. I cannot ignore what science has discovered concerning the functioning of my brain. I understand that my unreflected 'seeing' is naïve. It cannot be that I simply look out of my eyes and see the chair. It is a demonstrated fact that my seeing of the chair depends on processes occurring in my brain of which I have no immediate awareness.

And yet, we should pause and question what is happening here. In one moment I am speaking of being in a thought-free state, and in the next I am saying that such a state is inaccessible to ordinary reflection. If, as I claim, my knowledge of being in a thought-free state is not founded on acts of reflection, then how is it founded? For, if I truly remain in a thought-free state, then I do not think, I do not comment on what is occurring, I do not characterise it, my being present in the world simply is what it is. There is (inner) silence.

The answer here is best understood with reference to an ordinary act of perception. For example, I hear a noise in the garden and *think:* is that a cat? I go to the kitchen window and I *look* to see if there is a cat and then I *see* a cat. Although thought and reflection were involved in my first framing the idea that there could be a cat in the garden, neither my act of looking nor my act of seeing were acts of thought or reflection. I can look and see without thinking because I *already* have the garden immediately before me, i.e. it is not something I have thought up or imagined.

In the same way, if I stop thinking, then my pure unreflected sensory experience is *already* present. And yet this purity is *precarious,* for, unlike my perception of the garden, it *depends* on my not thinking. As soon as I begin to think, I overlay the pure sensory experience with my reflected knowledge and I lose the very domain I am attempting to discover.

Here we should pause again. Simply to follow a description of a thought-free state does not mean that one has entered such a state. It is easy to mistake the thought that refers to the state for the state itself. This is our normal way of proceeding. After all we cannot be expected to actually find a garden and a cat in order to follow the example of looking and seeing. And yet, when it comes to a state of not thinking, the situation is different. I cannot meaningfully think about not

thinking. Everything that is claimed about such a state has to be tested 'now' or one is simply failing to follow what is being said.

Accordingly, what it means to 'look' into a thought-free state can only be discovered from within that state 'now.' So let us enact an experiment. I have already claimed that in a state of thought-free experience I have no notion that I am having an 'experience.' To test this I must enter the very state I am, at first, only thinking of. In entering that state I find I can still hold a question in mind: I can *look* to see if there is any notion of my having an experience present in the experience. There is no movement of thought involved. I immediately 'see' that there is no such notion.

It is important to examine this carefully, to *dwell* on how it is that I can have knowledge of a thought-free state without thinking about it. To begin, in order to question this state I must frame a question as a proposition, in language, and this surely involves me in thinking about that question. However, that does not mean that the *act* of questioning is itself an act of thinking. The issue centres on whether I can understand meanings without thinking. The test is to stop thinking and *listen* to someone else talking. *Within* this state, I look to see whether I immediately understand what is being said without any movement of thought or reflection. As far as I am concerned, I find I do understand what is being said, immediately and directly. I cannot say *how* I understand, I simply *do* understand, just as, if I were to listen to someone speaking in an unfamiliar language, I would *not* understand. Because I can understand language without thinking, I can question pure experience by holding the question-meaning in mind (without thinking about it) and 'look' to see how things stand between my experience and the question-meaning. This looking is an *openness* to the experience that holds my existing thoughts and opinions at bay.

Such 'looking' and its relation to language and reflection is fundamental to our whole enquiry. Unless we have some means of questioning experience, we shall have nothing concrete to reflect on, for pure (thought-free) experience, of itself, does not reflect. The action of *directly* looking into a thought-free state provides us with an immediate (i.e. unmediated by reflection or inference) knowledge of that state. It is on the basis of this knowledge that we can begin to understand experience *explicitly.*

We are, therefore, from the beginning, making the validity of our enquiry absolutely dependent on the validity of our 'seeing' directly into a thought-free state. Such seeing is *foundational* because we are taking it to be capable of revealing the *truth* concerning that state. That does not mean I cannot be mistaken in any assertion I may make concerning what I am seeing. There is

always room for error in terms of the *language* I use, or I can *imagine* that I have entered into a thought-free state, or I can *presuppose* the answer to my question and fail to properly test the correspondence. But, as with all genuine enquiry, I can stand corrected by another whose 'seeing' has been more acute than my own.

The basic event of 'seeing' into experience is the correspondence of a pure intended meaning (the question I have in mind) and a pure experience. I register this correspondence in the same way as I register the correspondence of my 'seeing' that there is a chair in front of me when I am having the visual experience of a chair being in front of me. Both 'seeings' carry a warrant of self-certainty within themselves. This warrant is their being grounded in my immediate experience 'now.' Experience, in itself, is what it is, and so provides the necessary ground, i.e. my experience of there being a chair in front of me now just is an experience of there being a chair in front of me now, whether or not I am 'correctly' perceiving it, or dreaming, or hallucinating.

The difference between immediate perceptual 'seeing' and my 'seeing' into pure experience, is that, in framing questions of experience, I have abstracted myself into a framework that already understands experience *as* experience. It is from that place, and in relation to that pre-understanding, that I begin to question my immediate experience, and seek to discover general truths concerning the nature of that experience. In contrast, in my immediate perceptual experience, I *implicitly* know *within* the experience that there is a chair in front of me. I do not 'set up' a framework, I am already and involuntarily 'thrown' into the world.[3]

The fact that my questioning of experience is informed by a pre-existing framework of understanding is both necessary and problematical. And yet we must start somewhere, and our questions must be intelligible. That is why we are already working within the reflected understandings that we find deposited in the language of our culture. These understandings of experience as subjective, of our being conscious, of our having evolved, of time and space, of neuroscience, and so on, are the necessary background from which our questioning has to emerge. Our task is to *test* these understandings against the *evidence* of pure experience.

2.4 Direct Knowledge of Consciousness

Having now indicated what it means to 'look' into the state of pure (thought-free) experience and to 'see' its essential character, we are in a position to renew our enquiry into consciousness itself.

[3] Cf. (Heidegger, 1962/2008).

Our question is whether consciousness is something we can encounter *directly,* or whether, like the concept of experience, it is a distinction that can only be drawn on the basis of our *reflecting* on experience.

What we are attempting is to look *through* experience toward the encompassing consciousness that makes experience conscious in the first place – the unified totality within which each particular 'consciousness of' is known. We are looking to see whether there is a knowledge of what it is to be conscious inherent in a state of pure thought-free experience. The answer, if there is such direct knowledge, is the direct knowledge itself. To attempt to express such knowledge in language, to attempt to *think* such knowledge, is to have misunderstood it. Pure consciousness (if it is not a fiction) *exceeds* language. It is the space in which language *appears.* This is knowledge that *requires* us to have left the domain of thought and language to one side. That is what makes it *direct.* If I cannot do this, then no proposition in language is going to reach me concerning such knowledge. It will remain a speculation, a possibility. What it means to be conscious is known immediately in the very state of being conscious without thought. One just has to *look.* Here, the 'looking' and the corresponding 'seeing' *are* the demonstration of a direct knowledge of consciousness. This knowledge cannot be expressed in language because it is *too* immediate – there is no separation between the subject and the object. Consciousness is not an object. It is that by means of which objects are known. I cannot define knowledge of consciousness, and I do not *need* to define it. I define what is separate from me, what I am not, the entities I am 'conscious of.' But I am not 'conscious of' consciousness. 'I' am this consciousness. This is not difficult until you think about it.

It should now be clear that an ability to enter into a thought-free state is central to understanding the remainder of this thesis. Without such access there can be no direct knowledge of consciousness and without such knowledge much of what I have to say will appear groundless. However, I will not be basing the entire thesis on the direct evidence of my own experience. I will also be arguing and seeking to demonstrate that a series of analogous direct knowledge of consciousness realisations already illuminate the core of contemporary Western philosophy – beginning with Descartes and reaching explicit expression in the foundational works of 20th century phenomenology. My plan is to allow the thought-free state to act as a key to uncover the concealed history of this tradition, while at the same time using the core insights of the tradition to

illuminate what it means to realise a direct knowledge of consciousness. In this way I intend to *triangulate* what has already been said.

Chapter 3

Descartes

We begin the task of triangulating the material presented in the previous chapter with an investigation of Descartes' *Meditations.* That is not to suggest that no other philosopher before Descartes is worthy of a similar investigation. In particular, Plato's Myth of the Cave already presents an archetype of the philosophical task of escaping from a normal state of consciousness in order to attain to true philosophical knowledge. However, with Descartes, the parallels with what we have been saying become even more direct and definite. Firstly, Descartes does not explain himself in terms of an allegory, but, in presenting his method of doubt, provides an explicit series of steps aimed at putting his normal state of belief in the world out of play (as we are doing by stopping thinking). Secondly, in having secured what he believes to be a doubt-free foundation, Descartes attempts to find his way forward by means of *clear and distinct perceptions.* Similarly, in stopping thought, my direct 'looking' and 'seeing' into experience become the central means of apprehending truth. Thirdly, in attempting to doubt everything, Descartes arrives at what he considered to be the fundamental and indubitable truth of his own existence. As we will show, this indubitability is evidence that Descartes too had encountered a direct knowledge of consciousness, which he then (less successfully) attempted to bring to language.

3.1 The Method of Doubt

To make these parallels clear, we shall consider Descartes' *Meditations* in the *light* of a pure (thought-free) state of consciousness. Our interest is in Descartes' attempt to find an unshakeable ground of knowledge within himself. It is this *intent* that is significant. Rather than present and defend a position, Descartes sets about dismantling his own position, and in so doing provides a

demonstration by example that he invites his reader to follow. He does not want to prove anything (yet). He wants to indicate something he has *discovered* on the basis of his own meditative philosophical enquiry. His method of doubt is not an intellectual method, i.e. something to be merely read and considered in thought. He means that you and I also set about suspending any belief that we can in any way doubt. This is an experiment one has to perform, otherwise one does not *discover* whether there is indubitable knowledge (one merely acquires an opinion).

So, our intent is to listen to Descartes and not assume that we already know where he is going. To begin, he famously casts doubt on the existence of the world as it is known through the senses:

> At this moment it does indeed seem to me that it is with eyes awake that I am looking at this paper; that this head which I move is not asleep, that it is deliberately and of set purpose that I extend my hand and perceive it; what happens in sleep does not appear so clear nor so distinct as does all this. But in thinking over this I remind myself that on many occasions I have in sleep been deceived by similar illusions, and in dwelling carefully on this reflection I see so manifestly that there are no certain indications by which we may clearly distinguish wakefulness from sleep that I am lost in astonishment. And my astonishment is such that it is almost capable of persuading me that I now dream (Descartes, 1997, p. 19).

Here Descartes is questioning our belief that the things and events of our waking life have their foundation in an external world that exists independently of our experiencing those events. The point is that I cannot be certain that such a world exists in itself. Descartes is *astonished* by this – it has hit him with the force of a realisation – one that causes him to *actually* suspend his belief in the existence of the external world. Similarly, if we are to fully understand Descartes, we too must suspend belief in the external world, not as an intellectual 'as if' exercise, but as a foundational shift in our understanding of experience, one where we too have the sense that it indeed *could be true* that I dream while I am awake.

Descartes now goes down another level. For even if I consider the independent existence of the external world to be an illusion, I can still hold certain propositions to be true that I can demonstrate independently of my senses. For example, I can think of an ideal Euclidean triangle, and on the basis of its formal definition, I can conclude that its angles must *necessarily* add up to 180^{o}. In comprehending the triangle it seems I also comprehend the necessity of the relation between the angles. But even here, Descartes (the mathematician) want us to doubt:

> And, besides, as I sometimes imagine that others deceive themselves in the things which they think they know best, how do I know that I am not deceived every time that I add two and three, or count the sides of a square, or judge of things yet simpler, if anything simpler can be imagined (Descartes, 1997, p. 21)?

If we take Descartes seriously (and we are), then this passage should make us pause. His idea is that it is at least *possible* that an all powerful God (or an evil demon, or myself, unbeknown to myself) could manipulate my basic processes of judgment so that I am deceived even in those matters that lie entirely within the realm of my immediate experience. We shall clarify what lies in this realm as we continue – for now it at least includes intuitions of idealities such as number and geometrical shape. The crucial question for us is how far Descartes intended this doubt to reach.

On one reading, it could be that I am mistaken that two and three make five because I am asserting this on the basis of a demonstration that I incorrectly *remember*. Similarly, it could be that I incorrectly count the sides of a square because my *memory* errs concerning the number of sides I have already counted. In both these cases, the doubt concerns the veracity of any thought I may have concerning the past, or of any judgment I make whose truth relies on something that is not immediately present to me ('now').

On a second reading, it could be that my very judgment 'now,' concerning a state of affairs that I perceive 'now' could be in error, such as my placing a group of two counters on a table next to a group of three counters and my perceiving ('seeing') that there are, in front of me now, a group of two and a group of three and that their unity is a group of six. Or I could draw a square on the same table, and while looking at the square 'now' I could assert that the square has five sides, on the basis that I perceive 'now' that the square has five sides. In neither case have I *counted,* I have *perceived* how many elements there are.

The second doubt goes further than the first in allowing that my immediate perceptual judgments (my 'seeings') concerning the content of my experience (as experience) could be in error. If doubt goes this far then we descend into unintelligibility and absurdity, for we must then doubt that the words we use have any stable reference. For example, I may assert (and believe) that I see a table in front of me when really I am having a visual experience of a chair. Such a case could be understood (externally) as the misapplication of the word table when I 'should' have been using the word chair. However, if I also *believe* that the chair I am seeing is a table, e.g. I believe it has a hard, flat surface with four corners, even though my visual experience is of cushions, arms and

so on, then my experience of a coherent world has 'come apart.'

Such a 'coming apart' is indeed possible, as the neurological evidence attests.[1] Is this what Descartes intended? For if we doubt experience at this level, how are we to make any intelligible assertion concerning a ground of certain knowledge? Surely, every assertion has the form of a relation between the language used to make the assertion and the matter that is asserted? And if all truth is founded on this relation, i.e. on our certainty of a correspondence between statements and the state of affairs they assert, and we now doubt this (inner) certainty of correspondence, then how can any knowledge statement be immune from doubt?

3.2 The Circularity of Descartes' Doubt

At this point, if we consider the logical forms of Descartes' arguments, certain contradictions begin to emerge. Firstly, as is well-known, he *did* conclude that we have one item of indubitable knowledge, namely the knowledge that "I am, I exist":

> So that after having reflected well and carefully examined all things, we must come to the definite conclusion that this proposition: I am, I exist, is necessarily true each time that I pronounce it, or that I mentally conceive it (Descartes, 1997, p. 25).

Secondly, he was clear that "I ought no less carefully to withhold my assent from matters which are not entirely certain and indubitable than from those which appear to me manifestly to be false" (Descartes, 1997, p. 18). From this we can infer that Descartes' doubts concerning the validity of addition and counting should also cause him (and us) to suspend such mathematical knowledge. But, according to our earlier considerations, if I doubt that two and three make five, then I should also doubt that "I am, I exist," i.e. because I could be deceived in my use and understanding of language. And if I can be deceived in my use and understanding of language, then no linguistic proposition can be certain. Descartes himself realises this point part-way through the third meditation:

> I am constrained to confess that it is easy to Him [God], if He wishes it, to cause me to err, even in matters in which I believe myself to have the best evidence. And, on the other hand, always when I direct my attention to things which I believe myself to perceive very clearly, I am so persuaded of their truth that I let myself break out into

[1] For instance, consider Oliver Sacks' famous case of the man who mistook his wife for a hat (Sacks, 1985).

> words such as these: Let who will deceive me, He can never cause me to be nothing while I think that I am, or some day cause it to be true to say that I have never been, it being true now to say that I am, or that two and three make more or less than five, or any such thing in which I see a manifest contradiction.
>
> And, certainly, since I have no reason to believe that there is a God who is a deceiver, and as I have not yet satisfied myself that there is a God at all, the reason for doubt which depends on this opinion alone is very slight, and so to speak *metaphysical.* But in order to be able altogether to remove it, I must inquire whether there is a God as soon as the occasion presents itself; and if I find that there is a God, I must also inquire whether He may be a deceiver; *for without a knowledge of these two truths I do not see that I can ever be certain of anything* [my emphasis] (Descartes, 1997, p. 22).

Here Descartes creates a situation from which he cannot escape. For if nothing is certain until he can establish whether there is a God (and whether He is a deceiver), then *how* can such an establishment take place? The (short) answer is, for Descartes, it cannot. This is because he already assumes the only ground of certainty is God Himself and the arguments he uses to show that God exists rely on the certainty he attaches to his clear and distinct perceptions of the truth of particular propositions. But these clear and distinct perceptions only gain their certainty (for Descartes) if God exists and He is not a deceiver. So Descartes is caught in a circularity of doubt from which no proof or certainty can emerge.

3.3 Descartes' Discovery

At this point, given Descartes' method of doubt is logically unfit to provide the kind of foundation for faith and the sciences that he envisaged, we should ask what Descartes' *actually* achieved. If we stay with the logical form of his project, then we must conclude his theistic dualism is of little interest, aside from having motivated so many others to correct him. But our concern is not with the outer form of Cartesianism but with the *inner significance* of Descartes' project. The question is: did he escape from his normal everyday state of consciousness, and if so, what did he discover?

Descartes' doubt, in the first instance, enabled him to realise that his belief in the existence of the external world is, in fact, a belief, and not a certainly given truth. Although this belief is a

part of the fabric of experience, the method of doubt created a separation in Descartes between a pure subject of experience and the world existing 'in' and 'for' that subject. Descartes then sees it is possible (thinkable) that we dream while we are awake. In this way he *discovers* a world of immanent subjective experience. He also discovers that this *entire world* is immune from his preliminary doubting:

> Finally, I am the same who feels, that is to say, who perceives certain things, as by the organs of sense, since in truth I see light, I hear noise, I feel heat. But it will be said that these phenomena are false and that I am dreaming. Let it be so; still it is at least quite certain that it seems to me that I see light, that I hear noise and that I feel heat. *That cannot be false* [my emphasis] (Descartes, 1997, p. 29).

As Husserl is later to recognise, Descartes has uncovered the world of pure phenomena, i.e. purified of their reference to an independently existent external world. He is in the presence of 'what is,' of pure experience 'now,' of that which 'cannot be false.' Insofar as he does not add anything to this state (i.e. by thinking about it), he has left his normal everyday consciousness behind. Descartes also recognises that in order to maintain this state he must not revert back towards his former beliefs. Instead he must only trust to what he can clearly and distinctly *perceive:*

> I am certain that I am a thing which thinks; but do I not then likewise know what is requisite to render me certain of a truth? Certainly in this first knowledge there is nothing that assures me of its truth, excepting the *clear and distinct perception* of that which I state, which would not indeed suffice to assure me that what I say is true, if it could ever happen that a thing which I conceived so clearly and distinctly could be false; and accordingly it seems to me that already I can establish as a general rule that *all things which I perceive very clearly and very distinctly are true* [my emphasis] (Descartes, 1997, p. 35).

Here we must be careful to distinguish between Descartes' understanding of 'thinking' and our own. For Descartes, 'thinking' encompasses all possible acts of consciousness (including perceiving), whereas, for us, thinking and perceiving are entirely distinct ways of understanding. This is not just a terminological difference. As we shall see, it was Descartes' failure to understand this distinction that caused him to leave the world of pure phenomena behind and attempt to prove the existence of God. Nevertheless, I shall take it that Descartes' method of doubt has enabled

him to enter into a state of pure 'looking' and 'seeing' which he then describes in the language of clear and distinct perceptions. It is on this basis that he 'sees' that no pure phenomenon of experience can be false. It is also on this basis that he realises what he takes to be the fundamental and indubitable truth: "I am, I exist." It is in this realisation he first encounters what we have called a *direct knowledge of consciousness.*

The salient feature here is that Descartes held this truth of his own existence *above all others* (even above the truth of mathematical propositions) although he was unable to explain exactly why. On reflection he conceded that God could deceive him, but still wanted to protest: "Let who will deceive me, He can never cause me to be nothing while I think that I am." Here Descartes is on the cusp of realising that *in the moment I am conscious without reflection,* there is a direct knowledge of being conscious, of existing, that does not permit of doubt because it is not the knowledge of the truth of a proposition, it is a direct knowledge of *being now.* It is *this* direct (pre-propositional) knowledge of consciousness that cannot be doubted. In saying "I am, I exist" Descartes is only attempting to bring this knowledge to language.

However, as we know, language introduces doubt. Perhaps I am misled in my understanding of the meanings of the words I use. And who or what am 'I'? Perhaps my world has 'come apart' in some way that I do not realise, so that my self-understandings, memories and perceptions are unstable and confused. If so, is the *proposition* "I am, I exist" still certain? Descartes, in the end, and under the pressure of such doubts, had to admit he could not be certain. And yet, he remained split, unsettled and unconvinced.

The way out of this impasse is to recognise that even if my experience has become incoherent, *it is still the experience that it is.* 'I' cannot be deceived about unreflected experience because 'I' am not asserting anything concerning that experience. In not reflecting, 'I' am not separate from that experience. There *is* the experience, and the experience *is* conscious (otherwise there would be no experience). In this pure (thought-free) state, consciousness reveals (shows) *to itself* whatever it is that is being experienced 'now.' In this 'showing to itself' there is no *room* for doubt, for there is no separation of an 'I' that could reflect and doubt.

To be conscious without thought (reflection) is to know what it is to be conscious directly and immediately. One simply has to 'see' this is true, not as a timeless proposition, but by continual verification 'now' and 'now' and 'now.' Such knowledge is immune from doubt because it is *beyond* doubt. Once thought is relinquished, it is simply not *possible* to doubt what is revealed, because to doubt one needs to think. So, the indubitability of pure experience is a simple tau-

tology. And yet there remains (in each moment) the immediate accessibly of the knowledge of what it is to be conscious. This knowledge is utterly subjective – it cannot be verified by anyone else – and it cannot be known by the 'I' of everyday consciousness. Everyday consciousness is a 'consciousness of.' It 'knows' that there is a cat in the garden, that 'my' foot is itching, and so on. But it cannot encounter a direct knowledge of consciousness, it can only *think about* such knowledge.

3.4 Perception and Reflection

Descartes' fundamental oversight was not seeing the distinction between his clear and distinct perceptions and his (after the fact) reflections on experience. It is the thought-free state of pure or direct perception that remains immune from doubt. Any subsequent attempt to reflect on what is perceived – to conceive or think it – breaks the immediacy of the direct perception and allows the reflection to frame the experience according to the pre-understandings that structure the reflecting mind.

It should be noted here that in distinguishing between perception and reflection we are not saying that *concepts* have no role to play in perception. To address such a question would involve us in another kind of enquiry.[2] Instead, we are looking directly into experience and giving names to the distinctions we find. For us, the state of pure or direct perception is what remains when the otherwise continuous stream of inner thought and reflection is left behind. It is in *this* state that I first clearly and distinctly perceive the experiential world. Descartes does not fully understand the state because he already possessed a *theory* of perception:

> . . . when looking from a window and saying I see men who pass in the street, I really do not see them, but infer that what I see is men . . . And yet what do I see from the window but hats and coats which may cover automatic machines? Yet I judge these to be men. And similarly solely by the faculty of judgment which rests in my mind, I comprehend that which I believed I saw with my eyes (Descartes, 1997, p. 32).

Here Descartes *abandons* his immediate experience and begins to *theorise.* If we stay with experience itself, I *do* see men wearing hats and coats and I am aware of no act of (conscious) judgment. Therefore I can quite correctly say that I see men walking in the street. In contrast, I

[2] Such as McDowell's investigation of mind and world (McDowell, 1994b).

have to leave the state of pure perception and deliberately reflect on my experience before I can separate out the visual experience of hats and coats and *conceive* the possibility that there may be something other than men occupying those items.

Of course, the question is *how* can I say that I perceive men wearing hats and coats (rather than, say, nameless blobs of colour moving) and at the same time assert that perception precedes language? The brief answer (which will be elaborated later) is that my perception of there being men in the street *contains* an implicit *expectation* of how the perception will *unfold.*[3] For example, should a hat tilt backwards I expect to see a face. This does not require me to form an assertion of what is happening in order to understand what is happening. It is enough that events broadly unfold according to my (perceptual) expectations. This is not an ungrounded theory. It is my experience that I still perceive the world without thinking 'there is a cat,' 'there is a table,' and so on.

So Descartes, despite discovering a state of pure consciousness on the basis of his method of doubt, is almost immediately misled by a pre-existing understanding of perception as something separable from 'mere' sensory experience. Once he allows this theory to overrule his actual perception, the purity of his original insight is corrupted. He then begins his descent back into the world of his pre-existing opinions and beliefs, from which he unsuccessfully attempts to prove the existence of God.

Nevertheless, Descartes made a tremendous discovery. He found the *lumen naturale,* the natural light, which makes manifest the clear and distinct perception of truth:

> I cannot doubt that which the natural light causes me to believe to be true, as, for example, it has shown me that I am from the fact that I doubt, or other facts of the same kind (Descartes, 1997, p. 38).

For Descartes, the first and unassailable truth is the truth of the light itself. His expression "I am, I exist," his certainty that this cannot be doubted, is his realisation that consciousness 'is now.' What he fails to explicitly recognise is the extraordinary nature of this realisation and the state he has attained in order to reach it. As a result, rather than scrupulously maintain a state of pure perception via the complete suspension of all his former beliefs, he falls back into a state of thoughtful reflection and identifies the truth of his realisation with its assertability, and hence

[3] cf. Husserl's account of object perception (Husserl, 1997).

with language.[4] Once he reflects, Descartes can doubt his realisation, as any realisation can be doubted, because he can doubt his own use of language, and even his sanity. Now he needs God to underwrite his knowledge, because only if God is no deceiver can he be sure that his reflections are reflections of the truth. Nevertheless, and despite reasoning in this way, Descartes still attempts to prove that God exists on the basis of his clear and distinct perception of the truth of certain propositions. His *instincts* say that he sees the truth, even though his conscious reason has concluded otherwise.

3.5 The Problem of Conceptualisation

From the perspective we are taking here, Descartes is an *initiator.* He sets in motion the foundational inspiration of Western scientific culture: the idea that truth can be discovered on the basis of one's own experience (rather than on the basis of received authority). More than that, in his method of doubt, Descartes attempts to clarify the very ground of truth, and discover a fixed point of certainty upon which the entire edifice of Western science can be constructed. This ground is to be revealed through clear and distinct perceptions made manifest in the natural light of reason.

However, when we examine Descartes' actual philosophy, it becomes clear that the truth claims he makes on the basis of his clear and distinct perceptions do not possess the kind of certainty that he hoped for. This is because, in practice, Descartes was unable to distinguish between his clear and distinct perceptions of truths revealed in direct experience, and certain basic beliefs that so coloured his experience that he *mistook* them for perceptions of the truth. As an example, we need only remember his attempt to prove the existence of God:

> There is *no doubt* that those [ideas] which represent to me substances are something more, and contain so to speak more objective reality within them [that is to say, by representation participate in a higher degree of being or perfection] than those that simply represent modes or accidents; and that idea again by which I understand a supreme God, eternal, infinite, [immutable], omniscient, omnipotent, and Creator of all things which are outside of Himself, has *certainly* more objective reality in itself than those ideas by which finite substances are represented [my emphasis] (Descartes, 1997, p. 40).

[4] This identification is understandable, given we have to wait for the 20th century philosophy of Heidegger before the distinction between truth and assertability is made *explicit* (Heidegger, 2002).

Here the philosopher who previously took such pains to establish a single basic truth (I am, I exist) finds there is 'no doubt' that there is more 'objective reality' in ideas that represent 'substances' than in ideas which represent 'modes or accidents.' He fails to investigate the provenance of these concepts, and now freely asserts that all the premises of his argument are 'manifest by natural light.'

Nevertheless, by recognising the parallels between Descartes' method of doubt and our own investigation of consciousness, we can see that he was *on the right track.* He made the breakthrough into the domain of pure consciousness but was unable to distinguish the underlying essential structures of this domain from his pre-existing understandings. Consequently, he was unable to adequately *conceptualise* what he had discovered. The method of doubt allowed him through the door, so to speak, but once inside Descartes began to 'read into' his experience pre-existing conceptual distinctions that were not directly informed by that experience, but rather by the tradition in which Descartes was educated.

This is the problem of *conceptualisation.* To enter a state of pure unreflective perception is one thing – one has the immediate experience, it is what it is, it cannot be false.[5] However, as a philosopher, it is not enough simply to remain within the truth of my immediate experience. In order to speak of experience, I must bring my perceptions to language, and that means bringing them under objective, publicly understandable *concepts.* It is this procedure that is fraught with the possibility of error. For I continuously interpret my experience according to a certain framework of understanding that, in the first place, I acquired from the culture within which I grew to maturity. And it is not as if I could do without this framework of understanding – for without the framework I cannot make myself comprehensible.

What is needed is to bring my immediate experience to concepts in such a way that those concepts reflect the experience itself and not what I already believe that experience to be. This is not as difficult as it may first appear. For we already possess a natural ability to form concepts. How else did we, as children, acquire the concept of a dog, or a chair, or a table? It is not as if we had the nature of concept formation explained to us beforehand. We quite naturally learnt to

[5]That is not to say that my perception of a snake in *this* moment, cannot turn into a perception of a stick in a *later* moment. Once I perceive the stick, I re-evaluate my previous experience and take it to have been an illusion. In making that re-evaluation, I now consider my first perception to have falsely presented the objective (intersubjectively shared) world. Nevertheless, the initial experience, as an immediate experience, was still an experience of perceiving a snake, and, as such, was true *in itself* – i.e. without reference to an external reality outside of my immediate perceptual experience.

bring our perceptual experience under the requisite concepts. However, a child has an important advantage over the adult, in that (initially) the child lacks high-level pre-conceptions concerning what it is they are experiencing. They do not think about what they are doing, they simply 'look and see' what the various objects have in common that belong to the particular concept class that they are learning. In contrast, we, as reflecting adults, especially as philosophically educated adults, encounter the domain of direct experience already primed with an understanding of what it is that we experience. It is this understanding that interferes with and blocks our natural concept forming abilities.

So, the primary issue is not how to form adequate concepts of experience, but how to keep our existing pre-conceptions concerning the meaning and structure of experience out of play. If we can do that then our natural ability to form new concepts on the basis of pure observation is free to operate. Descartes doubted only to the point of realising that he could not doubt his immediate experience. And he recognised that the signature of this realisation was his clear and distinct perception of its truth. But he was unable to maintain his pure perception of immediate experience. Instead, he was misled by *feelings* of certainty concerning the truth of propositions that he was unable to ground in immediate experience.

Ironically, Descartes entertained the idea that there could be a deceiver working against him to subvert the truth, only to reject this on the basis of an argument that deceptively purported to prove there is a God who would not allow Descartes to be so deceived. In fact, Descartes is deceived by *himself,* by his own feeling of self-certainty, and by his lack of self-knowledge concerning the pre-conceptions he brought to bear on his own experience.

Chapter 4

Schopenhauer

The basic question that remains unasked in Descartes' *Meditations* is *who* exactly is the subject of the meditations? Descartes, once he has affirmed that God exists and that He is no deceiver, reverts to an orthodox understanding that *he,* Descartes, the man, the individual, is that subject of experience. He is a 'thinking thing,' a piece of unextended 'thinking substance,' that (presumably) remains distinct from the thinking substance of all other human individuals. And yet, who is Descartes, once he has doubted the existence of an external world and the existence of his body in that world? From such a place, he must also doubt the existence of any other human consciousness or 'soul.' His only certainty is the certainty of existing. It is from *this* place that the question again arises, who am I?

Of course, Descartes did *not* remain in the place of extreme doubt. As God is no deceiver, Descartes concluded that his perceptual understanding of the world (as being made up of extended bodies existing independently of his experience) must indeed be true. On this basis he developed his substance dualism of unextended thinking minds and extended unthinking matter. Then British empiricism, in the guise of Berkeley, undermined the notion that there need be any 'extended substance' whatsoever. For God does not deceive us about the world: it is a world of perception; to exist is to be perceived (with God acting as the universal perceiver). Beyond this there is no necessity to posit the existence of some unknowable extended substance that duplicates what is perceived. Hume then takes this early empiricism to its conclusion: if all we can know is what we experience, then we have no basis to conclude that there is any necessary connection between the things that we experience. Neither do we find anything resembling an enduring human self. What remains is experience experiencing itself, structured according to certain underlying principles of association.

Hume is then exceeded by Kant, who destroys the empirical notion of the human mind being a 'blank slate' on which experience is unproblematically written. Kant sees that experience can be divided into *that which appears* and the underlying universal *forms* of those appearances. His insight is that we cannot have *learnt* these underlying forms on the basis of having experienced them, as all experience presupposes these forms in order to be an experience of anything in the first place. Therefore, these forms (such as the forms of *space* and *time*) must in some way *precede* experience, and stand ready to give experience form. And yet this preceding cannot be a preceding in *time,* as time itself is one of the forms in question. It is rather that these forms stand *behind* experience as a kind of ground that makes experience possible.

Here we must be careful to distinguish between the *originary* spatiality and temporality of our senses and the *objectivity* of external space and time. For it is clear that we *do* learn to correctly perceive the objective space and time that we measure in terms of metres and seconds. But we do not *create* the originary spatiality and temporality of the sensory fields within which we perceive the objective dimensionality of the world. It is rather that a *primordial* sensory dimensionality stands ready to present our particular experience of there being an objective world, just as the sensory qualities (sound, colour, etc.) stand ready to fill out these sensory dimensionalities. For example, we do not learn that one thing follows another in *general,* we learn that *this* thing follows *that.* Our experience already has the form of a temporal *flow* that presents an immediate 'now' of experiential content. It is only within the framework of this flowing experiential 'now' that we can learn to distinguish particular things appearing in an objective (external) world.

Kant's reflections on the forms of experience led him to question what the 'things' we encounter in experience could be 'in themselves,' i.e. independently of their being experienced. Here he reached his famous conclusion that the thing-in-itself is in principle beyond our reach or comprehension, because we can only know of it in terms of the forms of experience, and these forms originate in *us* and not in the thing-in-itself. It is here that Schopenhauer takes up Kant's insights in his central work: *The World as Will and Representation.*

Schopenhauer stands out in this sequence of philosophical development because he opens up the question of *who* the subject of experience *is* at a level that *exceeds* Descartes. Armed with Kant's insights into the *a priori* forms of experience, Schopenhauer begins an investigation into 'things in themselves' that begins with his own direct experience. In so doing, he reaches again into the realm of the *cogito* that Descartes first uncovered in the extreme of his doubt. However, Schopenhauer now understands, as Descartes did not, the difference between perception

and reflection:

> As from the direct light of the sun to the borrowed light of the moon, so do we pass from the immediate representation of perception, which stands by itself and is its own warrant, to reflection, to the abstract, discursive concepts of reason, which have their whole content only from that knowledge of perception and in relation to it. As long as our attitude is one of pure perception, all is clear, firm, and certain. For there are neither questions nor doubts nor errors; we do not wish to go farther; we cannot go farther; we have rest in perceiving, and satisfaction in the present moment. Perception by itself is enough; therefore what has sprung purely from it and has remained true to it, like the genuine work of art, can never be false, nor can it be refuted through any passing of time, for it gives us not opinion, but *the thing itself.* [my emphasis] (Schopenhauer, 1969 i, p. 35).

Here Schopenhauer demonstrates a direct knowledge of the thought-free state of pure perception. What Descartes only glimpsed, Schopenhauer now opens up with conscious clarity. He sees again that pure perception cannot be false or doubted, but now recognises that all 'abstract, discursive concepts of reason ... have their whole content only from that knowledge of perception and in relation to it.' At one level, we can understand Schopenhauer as simply reiterating the empiricists' claim that all knowledge comes from experience. However, Schopenhauer goes further. He understands that the 'attitude' of pure perception is a distinct state of consciousness: 'we do not wish to go farther; we cannot go farther; we have rest in perceiving, and satisfaction in the present moment.' From within this state, which 'stands by itself,' it becomes clear what perception *excludes*, i.e. 'reflection' on the 'abstract, discursive concepts of reason.'

4.1 Concepts and Percepts

As we have already emphasised, it is one thing to enter into a state of pure thought-free perceptual consciousness, and it is another to conceptualise that state and bring it to language. It is the transition to language that remains fraught with the possibility of error, for, in attempting to *reflect* on experience, we leave the very state we wish to characterise and enter the state we had to suspend in the first place.

As soon as I start to *reflect,* I enter a network of thought structures whose origins remain obscure. I can usually and automatically find an interpretation of my experience, without in the

least being aware of the ground of that interpretation. It arrives ready-made, and presents itself as a true picture of the situation it depicts. I then understand my experience *through* the interpretation, and this understanding structures my experience in such a way that the interpretation appears to have its origins in the experience and not in my pre-suppositions concerning that experience.

Nevertheless, I am able to make basic assertions concerning my *perceptual* experience that remain unproblematically true to the experience – such as my saying there is a table in front of me now. Here, the state of affairs that I assert is grounded in my immediate experience. I directly *perceive* there is a table in front of me now. The question as to whether there 'really is' such a table, existing independently of my perceiving it, simply does not arise within the experience. To frame such a question, I must reflect on abstract concepts that are no longer grounded in the experience.

However, as a philosopher, I must move from making true assertions concerning particular perceptual experiences, to making true assertions concerning experience in general. The task is to continue to ground whatever is claimed back into direct perceptual experiences. Schopenhauer understands this task, but lacks any clear methodology whereby such a grounding can occur. This immediately leads him into difficulties with the distinction between concepts and percepts:

> Now although concepts are fundamentally different from representations of perception, they stand in a necessary relation to them, and without this they would be nothing. This relation consequently constitutes their whole nature and existence. Reflection is necessarily the *copy or repetition* of the originally presented world of perception, though a copy of a quite special kind in a completely heterogeneous material. Concepts, therefore, can quite appropriately be called representations of representations' [my emphasis](Schopenhauer, 1969 i, p. 40).

The problem is, despite their central role in all we have been saying, we have no clear, perceptually grounded idea of what it means for something to be a *concept.* For, if we remain with pure (thought-free) perceptual experience, we do not *directly* encounter anything that could be called a concept. What we encounter are tables and chairs, not the concepts of tables and chairs. Certainly, if we *reflect* on what it means for something to be a table, and for something to be a chair, we can see that these terms loosely define classes of objects according to certain properties that all the members of the respective classes have in common. But here we are simply recognising how certain language signs are related to our perceptual experience of the world. That does not show

us what a *concept* is in our direct experience, only that there is a certain structure in the way we employ language to refer to that experience.

If the concept of a concept is to be meaningful, then it must have its *origins* in experience, and this Schopenhauer admits. For him this origin is that of being a *copy.* But what is it that is being copied? If 'concepts are fundamentally different from the representations of perception' does this mean that there is no distinction between tables and chairs within my perceptual experience? And if there is such a distinction, does this not mean that my perceptual experience is *already* structured according to certain conceptual categories?

Here we must tread carefully, for we are leaving the immediacy of direct experience, and attempting to frame a concept that correctly captures the structure of that experience. We are asking whether my immediate perceptual experience is already structured according to the categories that I use to point out that experience in language. In other words, is it the case that I only know or realise there is a table in front of me when I reflect on this fact, or do I (implicitly) know there is a table in front of me *before* any such act of reflection has occurred? This again requires us to stop thinking and *look.*

To begin, even though I perceive this particular and unique table from this particular and unique vantage point, I nonetheless perceive it *as* a distinct unity, and not as an unidentified bundle of sensations. This 'perceiving as' imbues the perception with a sense of *recognition* that we ordinarily do not notice because we *continually* recognise what we perceive. It is only when we encounter something we do not recognise that our 'perceiving as' becomes salient. In that case we become curious, we investigate further, and so on.

So I can at least say that the table stands out in my pure perception as a distinct and familiar unity. I *implicitly* recognise it. This implicit recognition can be further demonstrated as I move around the table. I *expect* to see the partially obscured leg reveal itself as being attached to the table. I *expect* the shape of the table top to transform itself as I move in such a way as to maintain certain invariant relations. I am not directly aware of these expectations, because the focus of my conscious attention is on the table appearing 'now.' And yet these expectations contribute to my perceiving the *form* of the table. For I do not perceive the table as consisting only of the sides that are currently facing me, I take the table to have an enduring spatial shape, to be solid, to have a certain weight, and so on. Furthermore, if I interact with table, I find I know how to behave around the table, my actions demonstrate that I (implicitly) know what the table is *for,* that it affords my placing certain kinds of object on it, and so on.

In this way I see that a certain implicit idea or *concept* of the structure of the table is being fulfilled by the lawful flowing of my sensory experience. Such an experientially invoked concept, while structuring the experience, is, at the same time 'filled out' by it. The particular colours of the table present the table as the particular table that it is. I can see the unique detail of the table's shape and feel that shape confirmed as I run my hand along the table's edge. The experience *exceeds* the concept in its particularity and detail, while at the same time adhering to its inherent overall structure. The concept is *active* in the experience; it helps to form it; it shapes the thematic focus of the continuously changing stream of experience into a recognisably enduring table-structure that can only be decomposed afterwards, in an act of reflection.

Schopenhauer does not investigate his experience down to this level. Instead he takes over Hume's understanding that ideas (concepts) are mere copies of impressions (perceptions) (Hume, 1910, Section II). In so doing, Schopenhauer, like Descartes, allows an implicit *theory* of perception to overlay his immediate experience. Nevertheless, Schopenhauer's primary distinction between perception and reflection remains grounded in experience, because we can (and do) perceive without needing to reflect. However, the question remains, if reflection and perception are not distinguished by the presence or absence of concepts, then how are they distinguished?

Here we again turn to experience. The distinguishing feature of perception is that its conceptual content is filled out in a direct experience 'now' of a 'bodily present' particular. In reflection (in thinking) I leave my immediate experience and operate with concepts that are no longer embedded in the 'now' of a particular perception. I think 'about' my experience, I no longer 'inhabit' that experience – I abstract myself from the immediacy of experience 'now.' Furthermore, in immediate (thought-free) perceptual experience, only a select group of *concrete* concepts are actually invoked. These concepts are embedded in the experience and are only active insofar as the entities, events and situations they structure are themselves 'bodily' present. The distinction that Schopenhauer was attempting to draw between percepts and concepts is really a distinction between these concrete concepts undergoing immediate sensory fulfilment, and concepts (both concrete and abstract) that are entertained only in reflection. (Here, an abstract concept represents something that cannot be immediately encountered in a direct experience, such as the concept of an atom or a molecule).

What we see here is that direct perception does not place us in a state where concepts are inoperative. In not thinking, we do not thereby enter into a privileged state of direct knowledge of true objective reality. What we encounter is a direct knowledge of our own experience, unmedi-

ated by the interpretations of reflective thinking. In that experience we are already situated in the world, a world in which we are *embodied,* and a world that is already structured by concepts we have acquired on the basis of our *previous* experience.

Schopenhauer correctly realised that conceptual knowledge finally obtains its validity and content *from* direct experience. What he did not realise is that conceptual knowledge also *structures* that experience. He thereby failed to grasp that each act of consciousness contains or *intends* its own object. It is not until the work of Brentano that this intentional structure of experience is explicitly recognised (Brentano, 1995).[1] Schopenhauer's oversight here is understandable, but is also a symptom of a lack of methodological definiteness in his approach. Like Descartes, Schopenhauer proceeds according to an *instinct* for the truth. He has insight, and he realises the importance of grounding insight in direct experience. However, as his understanding of concepts and percepts demonstrates, he has no explicit way of keeping his existing philosophical pre-conceptions out of play.

4.2 The Concept of Will

Bearing the above considerations in mind, we now turn to Schopenhauer's fundamental philosophical idea: the identity of the thing-in-itself and the *Will.* For Schopenhauer, the immediate reality of the thing-in-itself is not unknowable, as Kant claimed, but *partially* reveals itself in our direct experience of willing. In contrast to the German Idealists (specifically Fichte, Schelling and Hegel), Schopenhauer does not take the 'purely knowing subject' to be the ultimate reality.[2] Instead, and in a way that brings Heidegger to mind, he starts with our immediate experience of being 'rooted' in the world:

> In fact, the meaning that I am looking for in the world that stands before me simply as my representation . . . could never be found if the investigator himself were nothing more than a purely knowing subject (a winged cherub without a body). But he himself is rooted in that world; and thus he finds himself in it as an *individual,* in other words, his knowledge, which is the conditional supporter of the whole world as

[1] We shall consider intentionality in more detail later.

[2] I should note here that the notion of a direct knowledge of consciousness has considerably more in common with the ideas of the German Idealists than it does with Schopenhauer's notion of the Will. For example, Fichte's absolute Ego points directly to the unity of an encompassing consciousness, and his opposition of the divisible Ego and divisible Non-Ego further points to the intentional structure of a 'consciousness of' the world (Fichte, 1889).

> representation, is nevertheless given entirely through the medium of a body, and the affections of this body are, as we have shown, the starting-point for the understanding in its perception of this world.

Here Schopenhauer observes that our direct perceptual experience is one of *embodiment.* It is the body, and the past experience that is embodied in it, that provides the context and structure within which our experience of the world unfolds. This is not the *extended* independently existing physical body that Descartes' doubted. It is the *experiential* body, the one I know from the *inside,* on the basis of direct experience. And it is through this direct experience that I encounter the true nature of bodily being:

> . . . the answer to the riddle [of the meaning of the world as representation] is given to the subject of knowledge appearing as individual, and this answer is given in the word *Will.* This and this alone gives him the key to his own phenomenon, reveals to him the significance and shows him the inner mechanism of his being, his actions, his movements. To the subject of knowing, who appears as an individual only through his identity with the body, this body is given in two entirely different ways. It is given in intelligent perception as representation, as an object among objects, liable to the laws of these objects. But it is also given in quite a different way, namely as what is known immediately to everyone, and is denoted by the word *will.* Every true act of his will is also at once and inevitably a movement of his body; he cannot actually will the act without at the same time being aware that it appears as a movement of the body. The act of the will and the action of the body are not two different states objectively known, connected by the bond of causality; they do not stand in the relation of cause and effect, but are one and the same thing, though given in two entirely different ways, first quite directly, and then in perception for the understanding (Schopenhauer, 1969 i, p. 100).

As the concept of the will outlined in this passage is fundamental to Schopenhauer's entire philosophical project, we should pause and *directly* test what is being said. Following Schopenhauer, we shall take 'true' acts of will to only be those that are immediately manifested as actions of the body. Every such act 'follows with strict necessity from the effect of the motive on the character' (Schopenhauer, 1969 i, p. 113). Furthermore, our willing is to be strictly distinguished

from 'resolutions of the will relating to the future' which are 'mere deliberations of reason about what will be willed at some time, not real acts of will' (Schopenhauer, 1969 i, p. 100).

Bearing these strictures in mind, let us assume I have the intention in me now to raise my hand in order to demonstrate to myself a particular act of will. The motive of demonstration is effective on my character and I find that my hand rises. For Schopenhauer, this action of raising my hand *is* my act of will, given outwardly as the perception of my hand rising and inwardly as my direct experiential knowledge of willing my hand to rise. However, if I stay with the actual experience, without thinking, I find the situation is not as Schopenhauer conceives it: I begin by entertaining (in reflection) the prospect of raising my hand in a way that I know has previously been effective in producing such an action, although I cannot describe exactly what it is that I do that makes the intention effective. I then await the result. After a pause I find that my hand rises. Or I can inwardly say to myself, 'I am going to raise my hand *now,*' and during the moment I inwardly say the word 'now,' I find my hand rises. But there is nothing in the experience to suggest that 'I,' the one who entertained the intention in reflection, had anything to do with the actual raising of my hand. As far as 'I' am concerned, it just happened; I know not why, and I know of no necessity that made it happen. I can equally imagine entertaining the same intention to raise my hand, for the same motivations, and finding that my hand does *not* rise. In the *actual moment* that my hand rises, I find I do *not* have any direct knowledge of my willing it to rise. My hand is simply rising.

I know, of course, on the basis of past experience, that an effective intention to stop my hand rising could operate at any moment. Here it appears that I at least *allow* my hand to rise by not forming an intention to stop it. But this assumes that I am somehow in control of the intentions that arise in me, that there is some *will* in me that wills the act to continue, whereas all I *actually* experience is the appearance of certain intentions and the appearance of acts that correspond to these intentions. *Why* and *how* these intentions are effective, and on what basis they have arisen, is something that is not revealed within the experience itself.

Once again, according to the immediate experience of raising my hand, in the moment it is rising, I find I have *no* direct knowledge of *willing* my hand to rise. I only have the immediate perceptual knowledge that it *is* rising. Clearly, such direct perceptual knowledge of my own bodily movements is given to me in a way that is quite distinct from the way that movements of other objects are given. But there is nothing in such perception that gives me a clue as to the inner nature of my being. I simply have a different and richer kind of perceptual knowledge concerning the movement and inner states of my body than I do of the movements of other bodies.

Nevertheless, I can and do say that 'I' raised my hand. I take *responsibility* for the act. In this taking responsibility, I identify myself as the agent, as the one who *carries out* the act. It is in this identification that the notion of my will gets its ground. For if 'I' am the performer of the act, then I *am* the will in action. If I look for the performer of this act in my direct experience, I find it is my *body.* What occurs is that 'I,' the conscious subject of experience, identify myself (my being) with the body that performs the act, and in this identification I understand myself to *be* the inner reality of the body: its living, experiencing core. In Schopenhauer's terms, I become the subject of experience *appearing as individual.* It is this subject that 'realises' the identity of the will and the body:

> The identity of the will and the body . . . is itself the most direct knowledge. If we do not apprehend and stick to it as such, in vain shall we expect to obtain it again in some indirect way as derived knowledge. It is a knowledge of quite a peculiar nature, whose truth therefore cannot really be brought under one of the four headings by which I have divided all truth, . . . it is the reference of a judgement to the relation that a representation of perception, namely the body, has to that which is not a representation at all, but is *toto genere* different therefrom, namely will. I should therefore like to distinguish this truth from every other, and call it *philosophical truth* (Schopenhauer, 1969 i, p. 102).

4.3 The Identity of Embodiment

Schopenhauer takes this direct knowledge of the identity of the will and the body to be the foundational insight on which his entire philosophy is to be based. He does not see that the subject of experience has identified *itself* with the body as it appears within experience. It is this identification that *creates* the notion of there being an individual will expressing itself through the individual acts of the body. If I *withdraw* from this identification (by ceasing all reflection) and become immediately present to experience as it is 'now,' then I find there is *no* direct knowledge of there being an individual will identified with an individual body. Conversely, if I *remain* within the identification, which I can only do in reflection, then my entire self-understanding is *conditioned* by the identification. I *become* an individual in the world because I *understand* myself as such. From *this* perspective, it appears to be a necessary truth that my inner being is the will that animates my body. But this truth is only necessary because I have already identified myself with

the body. I can immediately destroy this truth by withdrawing the identification.

What remains is pure experience. If I *reflect* on this experience, I can see it is an experience of embodiment. But it does not follow from the fact that I *experience* embodiment that I *am* a will that animates the body. What I *am* is given in a direct knowledge of consciousness. Such knowledge reveals my identity with the encompassing (impersonal) consciousness 'now.' It is *within* this consciousness that 'my' (personal) experience, my 'consciousness of' embodiment occurs. What Schopenhauer has realised in his direct knowledge of the identity of the will and the body is a direct knowledge of consciousness that has been *conditioned* by his *embodiment.*

The result of this misunderstanding is that Schopenhauer now conceives the true being-in-itself of the world according to his conception of the will as *personal.* Despite his realisation that all knowledge must be grounded in direct perception (and not in reflection), he has not seen how his understanding of perception is conditioned by a pre-existing understanding of embodiment. This pre-existing understanding immediately asserts itself once I reflect on experience. It is only on the basis of remaining with experience *as it is,* without reflection, by *looking* and *seeing,* that I can recognise the operation of this understanding, by recognising that it has *ceased* to condition my experience.

In suspending the conditioning, I once again find I have no direct knowledge of the action of a *personal* will, I simply experience *what happens.* Intentions arise in my reflective awareness – they may be effective, they may not. If I abstain from all reflection, then I find my body *moves by itself.* I discover that my *deliberate* intentions are not *necessary;* the body can stand up, walk, sit down, drink tea, and hold a conversation, all without my entertaining any intention to perform such actions. It *just happens.* Such direct knowledge reveals my identification with the body to be an *interpretation* of existence and not an immediately given truth. In ceasing to entertain intentions, 'I' as the initiator and executor of my actions have stepped out of the picture – 'I' do nothing, and yet the body carries on with its business. Within the identification, it 'feels' to me that 'I' am the one on whom the body's actions depend, that if 'I' do not intend the actions and keep on intending them by executing them, then the actions would not occur. But, in suspending all reflective activity, I put this reflected 'I' out of play, and then (to 'my' surprise) I find the body carries on quite well on its own.

4.4 The Feeling of Willing

From the perspective of direct experience, the 'I' of embodiment has only a relative or reflected existence. It is an artefact of the identification of impersonal consciousness with the experience of embodiment. Nevertheless, my embodiment is part of the *fabric* of my normal everyday understanding of being in the world. I do not *believe* I am embodied according to my personal preferences. It is something that is *presupposed.* It is there in my use of language and it is taken for granted in my interactions with everyone else. I can only recognise its functioning and bring it to distinct consciousness by withdrawing my otherwise unquestioned identification. Even then, my disidentification only lasts as long as I remain in a thought-free state, or as long as my reflective attention remains fixed on the investigation of such a state. Once the attention wanders, I am back in the everyday stream of consciousness that takes my embodiment for granted. I again *implicitly* understand myself to be *this* human individual, the inner reality of *this* human body. 'I' am the person who was born with (or in, or as) this body, who will die with it, and who 'holds sway' over it. 'I' am the author and executor of its actions, the thinker of its thoughts, and take for granted that 'I' am the source and origin of these thoughts and actions. I 'know' it and 'feel' it. And if I am unclear about what it means to have a personal human will, I only need examine what happens when it is *thwarted.*

Here we encounter perhaps the most compelling evidence that the human will has an independent reality: our experience of human *feelings.* For Schopenhauer, 'every impression on the body is also at once and directly an impression on the will. As such, it is called pain when it is contrary to the will, and gratification or pleasure when in accordance with the will' (Schopenhauer, 1969 i, p. 101). Again, Schopenhauer recognises an identity between something perceived (an impression on the body) and the will. This identity implies that when I experience pain, such as when I stub my toe, I also experience my will. The pain is my knowledge that this event of stubbing my toe is contrary to my will. However, if I look at an experience of pain directly, without reflection, I find it is just an experience of pain. It has a certain quality. It *hurts.* Taken in itself, it is a perception of a sensation located in my body. In the actual moment that the pain is experienced, it is what it is; there is no additional experience of it going against my will. Of course, the pain also indicates that the interaction that occasioned the pain may have injured my body and alerts me that some action may be required. So I look and see what occasioned the pain, and I look and see whether my toe is injured. These actions simply occur. I do not need to will them.

If I *reflect* on the experience, then I will say that I did not *like* the experience, that I did not *intend* the experience, that I would have avoided the experience if I could have. In that case I could say that the experience went against my will. But, once again, I see that my will only appears insofar as I reflect on the experience. My unreflected experience remains one of direct impersonal perception of sensations and actions.

But what of my *emotional* feelings, such as my *frustration* at having stubbed my toe? I am frustrated because I have been interrupted in what I was doing. I was late for an appointment. Now I am even later. The event has *got in my way,* it has frustrated my striving to arrive on time. Here my frustration is predicated on an understanding of myself that already assumes I am an embodied individual, that I have an identity that persists through time, that 'I,' the one who has stubbed my toe, am the same 'I' that will be late for my appointment. I project my 'self' into the future and I entertain intentions concerning how I want that future to be. If these intentions are blocked then I become frustrated. And yet I can only become frustrated insofar as I am *identified* with an intention and so believe myself to be the *intender* of the intention and to be the same self that will, at some future time, be in a state of having fulfilled or not fulfilled the intention. In engaging in such identification, I set up a situation of *wanting* the future to turn out a certain way. My intention, in intending a future, *persists.* It remains as a potential frustration that is triggered if its fulfilment is threatened. If it is triggered, it becomes *very difficult* to maintain a state of thought-free contemplation of pure experience. I am almost *forced* to reflect on the situation that the intention intended. It *demands* its right of entry into consciousness on the basis that I already gave it life by identifying with it in the first place. It is this experience of forced entry, and the subsequent feelings that arise on the basis of reflecting on the frustrated intention, that provide the most direct evidence of the existence of a personal will.

But, if we trace the phenomenon of this will back to its origins, we find again that it only *appears* to exist, only has its expression, on the basis of impersonal consciousness having already identified itself with my embodiment as a human self. This identification transforms an intention from something reflected in consciousness to something that is imbued with the living reality of my being. I *become* the human self that intends the intention. Once I am identified, I face the consequences. I know frustration, disappointment, success, relief, hatred, anger, jealousy, fear, and so on. And if I attempt to withdraw from the identification by suspending all reflective activity, I find I can only achieve this for as long as I am untroubled by my previous intending, i.e. so long as none of my intentions are being threatened. However, that does not mean a state of pure

thought-free contemplation is somehow *less* real than my emotional intending, or that, because my emotions can assert themselves over this state, they are somehow *more* real. The reason the emotions can assert themselves in this way is because the I of impersonal consciousness, in identifying with the activity of intending, has *allowed* this to happen. It nonetheless remains the case that if I abstain from all reflection, and remain entirely present with whatever is happening 'now,' no emotion, no *striving of will* can arise.

4.5 The Pure Subject of Knowledge

It is here that we find a basic contradiction emerging in Schopenhauer's thinking. He begins with the principle that all genuine conceptual knowledge must be grounded in direct perceptual experience and not in acts of reflection. He then asserts that our knowledge of the 'identity of the will and the body . . . is itself the most direct knowledge' (Schopenhauer, 1969 i, p. 102). And yet, according to our investigations, knowledge of my willing *only* arises on the basis of reflection. If we seek direct perceptual knowledge of the will, we simply find direct perceptions of the movements and sensations of the body. For Schopenhauer, everything hinges on there being a direct knowledge of the identity of the will and the body. This is knowledge he describes as having 'quite a peculiar nature,' that is known directly and *independently* of perception, and that so differs from other forms of knowledge that he calls it 'philosophical truth.' If *I* seek such knowledge, I do not find it. I only find direct knowledge of perception and direct knowledge of consciousness as the encompassing *container-enabler* of perception – knowledge not just of *what* I perceive, but knowledge *that* I perceive, that *I am,* that consciousness *is.* Once again, this must be *tested.*

What Schopenhauer *actually* discovered was not a truth concerning the identity of the will and the body but a truth concerning the structure of the *reflected* experience of being an embodied self – i.e. that such a state is brought about by impersonal consciousness identifying itself with its perception of a body. This identity is *not* an absolute philosophical truth because consciousness can *dis*-identify with the body – i.e. by abstaining from all reflection.

In short, Schopenhauer made a *mistake.* Like Descartes, he was unable to separate his immediate experience from his fundamental presuppositions concerning the nature of that experience (not because he was consciously biased but because he lacked a *disciplined* means of putting these presuppositions out of play). Then, as he was obliged to do, he took his discovery of the identity

of the will and the body to be the foundation of his entire philosophy. For him, this was the insight that set him apart from all those that went before him:

> I shall begin by producing a series of psychological facts proving first of all that in our own consciousness the *will* always appears as the primary and fundamental thing, and throughout asserts its pre-eminence over the intellect; that, on the other hand, the intellect generally turns out to be what is secondary, subordinate, and conditioned. This proof is the more necessary as all philosophers before me, from the first to the last, place the true and real inner nature or kernel of man in the *knowing* consciousness. Accordingly, they have conceived and explained the I, or in the case of many of them its transcendent hypostasis called soul, as primarily and essentially *knowing,* in fact *thinking,* and only in consequence of this, secondarily and derivatively, as *willing.* This extremely old, universal, and fundamental error, this colossal first false step, and fundamental confusion of ground with consequent, must first of all be set aside, and instead of it the true state of the case must be brought to perfectly distinct consciousness (Schopenhauer, 1969 ii, pp. 198–199).

At this point, we must ask, if Schopenhauer's philosophy is based on an error, then why have we chosen to consider him in such detail? Here we return to the question raised at the beginning of the chapter concerning the identity of the enquirer, of the one who *looks* and *sees.* Schopenhauer, in his investigation of the distinction between consciousness and the will has *opened up* this question by revealing the basic nature of our embodiment as a reflected experience of willing. It is by seeing into the nature of embodiment that we encounter the distinction between pure, disidentified, impersonal consciousness and my identified, embodied, personal consciousness of the world. The demonstration that Schopenhauer also saw this distinction is given in the following (remarkable) passage:

> Raised up by the power of the mind, we relinquish the ordinary way of considering things, and cease to follow . . . their relations to one another, whose final goal is always the relation to our own will. Thus we no longer consider the where, the when, the why, and the whither in things, but simply and solely the *what.* Further, we do not let abstract thought, the concepts of reason, take possession of our consciousness, but, instead of all this, devote the whole power of our mind to perception, sink ourselves completely therein, and let our whole consciousness be filled by the calm

> contemplation of the natural object actually present. ... We *lose* ourselves entirely in this object, to use a pregnant expression; in other words, we forget our individuality, our will, and continue to exist only as pure subject, as clear mirror of the object, so that it is as if the object alone existed without anyone to perceive it, and thus we are no longer able to separate the perceiver from the perception, but the two have become one, since the entire consciousness is filled and occupied by a single image of perception. ... Thus ... the person who is involved in this perception is no longer an individual, for in such a perception the individual has lost himself; he is *pure* will-less, painless, timeless *subject of knowledge* (Schopenhauer, 1969 i, pp. 178–179).

Here we find Schopenhauer describing just that state of impersonal consciousness that is present in all acts of direct perception. The difference for Schopenhauer is that he realises this state on the basis of a complete identification of himself as subject of knowing with a particular object of perception. He thereby encounters a state of impersonal ('will-less, painless, timeless') consciousness, but he does not turn the gaze of this consciousness onto his immediate experience. For him the key, the means of access, is *aesthetic contemplation.*

It is Schopenhauer's knowledge of this state that brings him to our attention. Just as Descartes escaped the everyday state of identification with his worldly consciousness through the method of doubt, so Schopenhauer escapes through aesthetic contemplation. He now recognises that 'the person who is involved in this perception is no longer an individual.' The *will* – that which was the 'primary and fundamental thing ... the true and real inner nature or kernel of man' – has been *silenced,* and the knowing consciousness, now will-less and impersonal, *remains.*

To enter this state Schopenhauer has explicitly stopped thinking ('we do not let abstract thought, the concepts of reason take possession of our consciousness') and disidentified with his human self ('we relinquish the ordinary way of considering things, and cease to follow ... their relations to one another, whose final goal is always the relation to our own will ... we no longer consider the where, the when, the why, and the whither in things ... we forget our individuality, our will'). He now has direct knowledge of consciousness. *This* is the *true* philosophical knowledge – the knowledge that *ought* to have shown him the deeper reality of his concept of the identity of will and body. For now, from this place of impersonal consciousness, Schopenhauer is no longer the individual, no longer identified with a body. In his own words, he is the pure subject of knowledge.

However, Schopenhauer is unable to bring this direct knowledge of consciousness back into suitable concepts, because he is already committed to an understanding of the will as thing-in-itself. Despite transcending his identification with individuality, he remains convinced that his insight into the world as will is correct. Here we can only speculate on Schopenhauer's deeper and unrecognised motivations. Firstly, there is the genuine insight that *ordinary* individual, knowing consciousness *is* conditioned by the identification of impersonal consciousness with the perception of a body. On the basis of this identification, when I *reflect,* I identify with what I *think* and with what I *intend* – I believe that *I* am the one who determines what I intend and that *I* am the one who is in control of this body. *This* is the error of my knowing consciousness that Schopenhauer exposes. For clearly most of my intending is determined by motives of which I am unaware. It is in this relative sense that we can say that the will (as that which determines my intentions) is primary and my knowing consciousness is secondary.

However, once I withdraw from my everyday state of reflective consciousness, I no longer identify myself as the intender of the intentions that arise in my reflective consciousness. My *personal* will is now abolished; it becomes something believed in on the basis of an identification that is no longer in operation. The impersonal consciousness that is revealed is *unconditioned* by my intending (i.e. by my personal will). I simply experience what occurs. The entire notion of my possessing a personal will is now seen as a kind of prism through which I experience reflective embodiment. From the place of impersonal consciousness it no longer makes sense to say that the will is primary and consciousness is secondary. If one had to form a concept of the will that corresponds with this state, it would be an impersonal will that simply wills whatever occurs. Such a will is neither primary nor secondary to consciousness. It would be better to say that it *is* consciousness, or that the impersonal will and impersonal consciousness are identical.

For Schopenhauer, such an understanding is not to be countenanced because it goes against his fundamental notion of the will as thing-in-itself. This is the distinguishing feature of his philosophy and the means by which he plans to save Kant's thing-in-itself from the absolute idealism of Hegel, and thereby change the course of Western philosophy. But Schopenhauer's realisation of the pure subject of knowledge presents him with the problem of explaining how such a state can come about if the will is primary and consciousness is secondary and conditioned by the will. To answer this, he develops the problematic idea that the will *denies itself:*

> But the man who sees through the *principium individuationis,* and recognizes the true nature of things-in-themselves, and thus the whole, . . . sees himself in all places

> simultaneously, and withdraws. His will turns about; it no longer affirms its own inner nature, mirrored in the phenomenon, but denies it. The phenomenon by which this becomes manifest is the transition from virtue to *asceticism.* In other words, it is no longer enough for him to love others like himself, but there arises in him a strong aversion to the inner nature whose expression is his own phenomenon, to the will-to-live, the kernel and essence of that world recognized as full of misery. He therefore renounces precisely this inner nature, which appears in him and is expressed already by his body, and his action gives the lie to his phenomenon, and appears in open contradiction thereto. Essentially nothing but phenomenon of the will, he ceases to will anything, guards against attaching his will to anything, tries to establish firmly in himself the greatest indifference to all things. His body, healthy and strong, expresses the sexual impulse through his genitals, but he denies the will, and gives the lie to the body; he desires no sexual satisfaction on any condition. Voluntary and complete chastity is the first step in asceticism or the denial of the will to live. It thereby denies the affirmation of the will which goes beyond the individual life, and thus announces that the will, whose phenomenon is the body, ceases with the life of this body (Schopenhauer, 1969 i, p. 380).

Here we see the immediate knowledge that I, as pure impersonal consciousness, am *not* my will, and therefore that I am *already free,* remains closed to Schopenhauer. His concept of the will as thing-in-itself keeps him attached to the *phenomenon* of the will by granting it a reality it does not possess (in itself). If the will were *real* in this way (i.e. if it lay beyond the forms of our knowing) then it could not be a phenomenon constituted in reflective consciousness. Hence impersonal consciousness could not simply 'see through' the concept of will – instead the will, as thing-in-itself, must somehow be the agent of its own denial. This is what leads Schopenhauer to his endorsement of asceticism:

> Just as he mortifies the will itself, so does he mortify its visibility, its objectivity, the body. He nourishes it sparingly, lest its vigorous flourishing and thriving should animate afresh and excite more strongly the will, of which it is the mere expression and mirror. Thus he resorts to fasting, and even to self-castigation and self-torture, in order that, by constant privation and suffering, he may more and more break down and kill the will he recognizes and abhors as the source of his own suffering existence

> and of the world's. Finally, if death comes, which breaks up the phenomenon of this will, the essence of such will have long since expired through free denial of itself except for the feeble residue which appears as the vitality of this body, then it is most welcome, and is cheerfully accepted as a longed-for deliverance (Schopenhauer, 1969 i, p. 382).

It is in this endorsement of asceticism that Schopenhauer's misunderstanding of the will begins to develop its consequences. He now believes his doctrine of the will receives indirect confirmation in the ascetic traditions of the various world religions (without himself having seriously engaged in ascetic practices). But the question concerning asceticism is *who* is the one who is castigating the body? Is this the action of the will-less and timeless subject of knowledge, or is it the action of the human individual, the one who is identified as the actor and intender? For the ascetic, the will is seen as something to be 'more and more' broken down and killed, and yet as something that cannot finally be destroyed as long as its 'feeble residue,' the will's phenomenon (the body), survives. Hence one must remain in a state of continual vigilance and self-conflict, unable to rest for fear the will will reassert itself:

> However, we must not imagine that, after the denial of the will-to-live has once appeared through knowledge that has become a quieter of the will, such denial no longer waivers or falters, and that we can rest on it as on an inherited property. On the contrary it must always be achieved afresh by constant struggle. For as the body is the will itself only in the form of objectivity, or as phenomenon in the world as representation, that whole will-to-live exists potentially so long as the body lives, and is always striving to reach actuality and to burn afresh with all its intensity (Schopenhauer, 1969 i, p. 391).

This is a picture of a *divided* self, where the will-to-deny-the-will must assert itself over the will-to-live. Schopenhauer's idea is that the pure subject of knowledge, having attained to a state that is no longer individual, sees the futility of all willing, and that this knowledge now acts as a motive for the will to deny itself. But throughout, Schopenhauer (and the ascetic) grants *reality* to the will. It is 'always striving to reach actuality and to burn afresh with all its intensity.' And yet, this will that attempts to deny the will-to-live, is the *same* will, now divided, now striving for what it conceives to be the greatest prize: freedom from itself, from all suffering, and the final attainment of bliss and union.

If I go back to my direct, unreflected experience, I find again that there is no willing or denial of willing occurring. I do not need to deny my will, I cannot deny my will, there is no one there to enact the denial. In order to deny an impulse of the will (i.e. an intention to act), I first must first *reflect* on what it is that is intended. I must then *reflect* on an already formed *resolution* to deny certain intentions that arise and *await* the forming of a counter-intention that I presume will arise on the basis of my having formed a resolution to deny the will. This entire sequence of intendings can only occur on the basis of my already having left the state of pure (selfless) unreflective consciousness. If I stay in this state, and an impulse arises, then I simply look and see whether the body is going to act on that impulse. There is no affirmation and no denial.

The entire drama of the denial of the will occurs in a reflective consciousness that is identified with the will. It *affirms* this state of identification, and thereby *denies* the state of pure consciousness, and yet it intends to attain a state of pure consciousness on the basis of this denial. As a method, I suspect it only has success because the extreme suffering that is induced finally causes the suffering self to give up, to become inactive, so that a state of pure consciousness is momentarily realised. But such a realisation is quickly lost because it is attained without a proper understanding of what is occurring. Instead the method of denial is itself affirmed because it is understood to have *produced* a momentary release. And so the cycle repeats itself.

The irony here is that Schopenhauer did not attain to a knowledge of pure consciousness by means of ascetic practice (despite insisting that all conceptual knowledge must be grounded in direct experience). His route, as we have shown, was through philosophical and aesthetic contemplation. In other words, his own experience was not one of a constant denial of the will. He was led by *inspiration* (i.e. by a love of truth and beauty) to a direct realisation of pure consciousness. It was this realisation that gave Schopenhauer's philosophy its power and insight. But, as with Descartes, the purity of Schopenhauer's direct knowledge could not be reflected in conceptual consciousness without becoming distorted. What was lacking, and what the entire discipline of philosophy was lacking, was a reliable means of recognising and eliminating such distorting preconceptions.

Chapter 5

Husserl

The question of developing a means of undistorted philosophical access to a state of pure consciousness leads us directly to the phenomenology of Edmund Husserl. It is here that we first encounter a philosopher who explicitly understands that the possibility of true philosophy is entirely dependent on the discovery and maintenance of such undistorted access. Husserl's route was by means of what he came to call the *phenomenological reduction.*

Viewed from a sufficient distance, the phenomenological reduction, like our own experiment in stopping thinking, is intended to bring the meditating philosopher into a state of pure consciousness. For Husserl, this state is not an end in itself, but is rather a domain to be secured in order to be explicated by means of phenomenologically purified *reflection.*

5.1 The Life-World and the Epochē of the Objective Sciences

Husserl developed various "ways" into phenomenology from his first "breakthrough" in the *Logical Investigations* to his culminating enquiry into the "pregiven life-world" in the *Crisis* text. This *Crisis* enquiry begins with a consideration of the origins of modern objective science which Husserl traces back to the development of Galilean mathematical physics. His aim is to demonstrate that our modern scientific understanding of the "true" (i.e. objective) being of the universe is founded on a mathematical idealisation of the world that is already given in our normal everyday experience:

> The contrast between the subjectivity of the life-world and the "objective," the "true" world, lies in the fact that the latter is a theoretical-logical substruction, the substruction of something that is in principle not perceivable, in principle not experienceable

> in its own proper being, whereas the subjective, in the life-world, is distinguished in all respects precisely by its being actually experienceable (Husserl, 1970/1992, p.).

It is this subjective-relative experienceable world that Husserl terms the "pregiven life-world" and it is the universal form of this life-world that he takes to be the proper subject matter of phenomenology. In explicating the a priori forms of this world, Husserl also seeks to provide a rational ground for the objective sciences by showing how the "truths" of science presuppose and depend on verifications that can only be obtained on the basis of life-world experience:

> The life-world is a realm of original self-evidences. That which is self-evidently given is, in perception, experienced as "the thing itself," in immediate presence, or, in memory, remembered as the thing itself; and every other manner of intuition is a presentification of the thing itself. [. . .] All conceivable verification leads back to these modes of self-evidence because the "thing itself" (in the particular mode) lies in these intuitions themselves as that which is actually, intersubjectively experienceable and verifiable and is not a substruction of thought; whereas such a substruction, insofar as it makes a claim to truth, can have actual truth only by being related back to such self-evidences (Husserl, 1970/1992, pp. 127-128).

Husserl's goal in investigating the Galilean origins of objective science is to show how our unreflective acceptance of modern science's mathematically idealised conception of the "true being" of nature causes us to overlook, or rather to *look through* the life-world in such a way that we no longer recognise it or distinguish it from our scientific model of the world. In order to remove this distorting layer of interpretation, and thereby to bring the life-world into clear view, Husserl requires that we each, as meditating philosophers, perform an epochē or bracketing of all objective sciences:

> This means not merely an abstraction from them, such as an imaginary transformation, in thought, of present human existence, such that no science appeared in the picture. What is meant is rather an epochē of all participation in the cognitions of the objective sciences, an epochē of any critical position-taking which is interested in either their truth or falsity, even any position on their guiding idea of an objective knowledge of the world (Husserl, 1970/1992, p. 135).

However, for anyone who reads the *Crisis* text it becomes clear that this epochē can only be effective if one has already understood what Husserl means by the objective sciences and the life-world that is to be distinguished from them. The enacting of the epochē is the bringing into play of this pre-existing understanding. The understanding itself relies on our careful reading of the *Crisis* text or on our having already and independently realised the distinction that Husserl is pointing out. From the perspective of the thesis, it turns out we *have* already discovered the life-world that Husserl is indicating: *it is the world as revealed in a state of pure pre-reflective (thought-free) consciousness.*

For us, therefore, there is no need to specifically effect an epochē of the objective sciences, as we can only (knowingly) bring scientific understandings into play by leaving (dividing) the state of pure consciousness and attempting to reflect on that state. If it were to be countered that we *unknowingly* bring scientific understandings into play that structure our *pre-reflective* experience, the answer is that such understandings would anyway belong to the structure of the life-world. Here we must distinguish between the particular life-world of a given individual or of a given historical period and the a priori structure of any life-world whatsoever. For Husserl, the idea of the first epochē is not to immediately uncover such a universal a priori structure, but only to bring into view my particular subjective-relative life-world.

5.2 The Phenomenological Reduction

The epochē of the objective sciences is therefore only a *first step* in effecting a phenomenological reduction. Husserl now wants to bring into question the world as it is understood quite apart from our scientific notions of objective existence. He wishes to question the very "pregivenness" of the life-world, our natural acceptance of its validity, of its actuality, our acceptance that we live *within* the world and that the world *precedes* us.

As with the epochē of the objective sciences, the intent is not to *deny* the validity of the actuality of the world and of our existence within it, but to suspend or bracket our normally unquestioned acceptance of these validities, in order that they may come into view *as* something we accept. Achieving this *transcendental* epochē requires more than a series of abstentions from individual validities:

> Instead [...] a completely different sort of epochē is possible, namely, one which puts out of action, with one blow, the total performance running through the whole

> of natural world-life and through the whole network (whether concealed or open) of validities – precisely that total performance which, as the coherent "natural attitude," makes up "simple" "straightforward" ongoing life. Through the abstention which inhibits this whole hitherto unbroken way of life a complete transformation of all of life is attained, a thoroughly new way of life. An attitude is arrived at which is *above* the pregivenness of the validity of the world, *above* the infinite complex whereby, in concealment, the world's validities are always founded on other validities, *above* the whole manifold but synthetically unified flow in which the world has and forever attains anew its content of meaning and ontic validity. In other words, we thus have an attitude *above* the universal conscious life (both individual-subjective and inter-subjective) through which the world is "there" for those naïvely absorbed in ongoing life, as unquestionably present, as the universe of what is there, as the field of all acquired and newly established life interests. They are all put out of action in advance by the epochē, and with them the whole natural ongoing life which is directed toward the actualities of "the" world (Husserl, 1970/1992, p. 150).

To attain this complete transformation, the individual:

> ... simply forbids himself – as a philosopher, in the uniqueness of the direction of his interest – to continue the whole natural performance of his world-life; that is, he forbids himself to ask questions which rest upon the ground of the world at hand, questions of being, questions of value, practical questions, questions about being or not-being, about being valuable, being useful, being beautiful, being good, etc. All natural interests are put out of play. [...] This is not a "view," an "interpretation" bestowed upon the world. Every view about ..., every opinion about "the" world, has its ground in the pregiven world. It is from this very ground that I have freed myself through the epochē; I stand *above* the world, which has now become for me, in a quite peculiar sense, a *phenomenon* (Husserl, 1970/1992, p. 152).

According to the path taken in the *Crisis* text, the transcendental epochē also requires that we develop a "new universal direction of interest":

> ... let us establish a consistent universal interest into the "how" of the manners of givenness and in the *onta* themselves, not straightforwardly but rather as objects in

> respect of their "how" – that is, with our interest exclusively and constantly directed toward *how,* throughout the alteration of relative validities, subjective appearances, and opinions, the coherent, universal validity *world – the* world – comes into being for us; how, that is, there arises in us the constant consciousness of universal existence, of the universal horizon, of real, actually existing objects, each of which we are conscious of only through the alterations of our relative conceptions of it, of its manners of appearing, its modes of validity, even when we are conscious of it in particularity as something simply being there (Husserl, 1970/1992, pp. 144–145).

Husserl's intention is that we bracket our natural acceptance of the actuality of the world and of ourselves as individuals existing within the world. This enables us to look into the manners of givenness of the "onta" that we ordinarily do not register because our focus of interest is on the actuality of things *as* the things they are in the world. In suspending acceptance of the actuality of the world, we come to recognise that our experience of the world is not a straightforward reception of something pregiven, but that actuality is something *bestowed* on experiential entities according to the how of their manners givenness in consciousness.

For example, I see a snake in front of me on the path, and in that moment it *is* a snake as far as I am concerned; it has attained acceptance by me, it is actual. I move a little closer and I now see it is a stick. The previous attribution of the actuality of the snake is cancelled and I now (involuntarily) find actuality attributed to my stick experience. The *reason* I no longer see a snake is because, when I approached closer, the harmonious, flowing, moment to moment transformation of my sensory experience (the how of the subjective manners of givenness) no longer matched my perceptual snake expectations.

The idea of the phenomenological reduction is that I now recognise my *entire* experience of the world stands upon a complex, many-layered and unified attribution of world-actuality. Ordinarily, in an experience of perceptual illusion, I remain within the "natural attitude" that takes the on-going background actuality of the world for granted. I do not see the attribution of actuality operating; I interpret my perceiving a snake as a *mis*-perception, an aberration, something out-of-the-ordinary, which occurred because I did not pay sufficient attention. However, such an experience, when reflected upon, provides a clue, a glimpse behind the curtain that reveals the subjective origin of actuality-attribution. The phenomenological reduction extends such an insight to encompass my total experience of the actuality of the world, including the actuality of my human self as an individual existing within the world. It is not that actuality-attribution ceases to

function, it is rather that I no longer *go along* with it in the same unrecognised way. I now *realise* that such actuality-attribution is occurring – my *seeing* of it means it can no longer operate invisibly, as if it were a property of the world itself. Instead I see how actuality (validity) is bestowed upon experience according to the manner of its harmonious unfoldment. In this way my normal experience of being in the world is *reduced* and I encounter the "transcendental subjectivity" that constitutes the meaning and actuality of my being in the world:

> . . . in the actualizing of the reduction a self-reflection occurs that has a wholly new kind of structure: it is not that man reflectively thinks about himself, but rather that transcendental subjectivity concealed in self-objectification as man, reflectively thinks about itself, beginning *seemingly* as man, annulling itself as man, and taking itself down as man all the way to the ground, namely, down to the innermost ground of its life (Fink, 1995, p. 32).

For Husserl, this seeing is not some kind of intellectual manipulation whereby I reframe my experience on the basis of a different set of assumptions; it is a profound discovery that creates a fundamental and lasting shift in the basic stance of my experiential consciousness:

> Perhaps it will even become manifest that the total phenomenological attitude and the epochē belonging to it are destined in essence to effect, at first, a complete personal transformation, comparable in the beginning to a religious conversion (Husserl, 1970/1992, p. 137).

In introducing the phenomenological reduction, Husserl is expressing something that cannot be grasped within a normal state of acceptance of the actuality of the world. The transcendental epochē is intended to put this acceptance out of play ("with one blow"). And yet, if we consider Husserl's explicit statements, the enactment of the epochē is a relatively straightforward matter of *not* taking an interest "in the being, actuality, or nonbeing of the world" (Husserl, 1970/1992, p. 175). It would appear that such an abstention can be easily achieved by means of a *thought experiment* that only grants actuality to the immanent stream of my own conscious experience. All else, the question of the reality of any transcendent world existing beyond the domain of my immanent experience, is put aside, and neither accepted nor rejected. I remain focussed on just that immanent stream and on what can be discovered within it.

If this is all Husserl means, then the transcendental epochē should lie within the reach of any philosophically educated individual who cares to seriously read the *Crisis* text. And yet, I predict

that the enactment of the epochē as a thought experiment will not, in itself, lead to "a complete personal transformation, comparable in the beginning to a religious conversion." Husserl intends more than a *hypothetical* suspension of belief in the actuality of the world – he means for that acceptance to *actually* be suspended. It is the meaning of this *actual* suspension that cannot be expressed within the domain of normal philosophical discourse because that discourse already assumes the actuality of the world. This actuality-acceptance *precedes* any hypothetical state of affairs I may construct, and remains in place during my intellectual entertainment of the possible actuality of that state of affairs. It is like a child's game, I 'make-believe' that my immanent stream of experience is the only reality, while (in reality) I know full well that it is not.

For these reasons, Husserl has to *indirectly* invite his reader into the reduction. To achieve this he leads us along paths he has already taken and speaks of the reduction from *within* the reduction in the hope that we may enact the epochē *empathetically,* perhaps as a recognition of something we already knew *implicitly.* From this place he describes what has been achieved in terms that can only be understood by someone who has *already* enacted the reduction. And yet, the actual *act,* the complete suspension of the "natural attitude," is not something that can be achieved from *within* the natural attitude. For if I understand what it is to be in the natural attitude, I have already transcended it. I am caught in a circularity: it is only by transcending the natural attitude that I can come to see that I have been caught within it, and it is only on the basis of such seeing that I can first transcend it. The act of seeing and transcending are one and the same. It is an immediate insight and not something that can be achieved in a series of rational steps. Husserl's collaborator and assistant, Eugen Fink, describes the situation as follows:

> Man's self-reflection first becomes a way into the transcendental attitude when it is "radicalised" in a sense such as is not possible in the natural attitude, radicalized, namely, to the *annulment* of the natural attitude. [. . .]
>
> In view of this situation, is there still any sense in speaking of *ways* into the transcendental attitude? If we take ways into phenomenology to mean a *continuity in motivation* that begins in the natural attitude and by inferential force leads into the transcendental attitude, then *there are no such ways.* That does not imply, however, that talk of "ways" into phenomenology is altogether senseless. Thus, for example, to start out from the Idea of radical self-reflection is one actual way, for in the performance of self-reflection of this kind there can spring up that transcendental il-

> lumination that first opens up the course of a self-reflection that has to be radicalized in a new sense; because on the occasion of a decisive and unwavering turn inward into oneself the dispositional possibility is created for catching sight, in a productive, anticipatory way, of the dimension of transcendental radicality. The way [into phenomenology] only becomes compelling if we already bring a transcendental knowing with us – even if one that is quite obscure (Fink, 1995, pp. 32–34).

Here, the entire success of the phenomenological reduction depends upon "a decisive and unwavering turn inward into oneself" on the basis of which I may (possibly) catch sight "of the dimension of transcendental radicality." In terms of our ongoing investigation into consciousness, this means I must already have discovered, according to my own efforts and power of insight, what it is to have transcended my ordinary, everyday state of consciousness. Only then may it become clear what Husserl is attempting to indicate by means of the reduction.

5.3 The Reduction and Direct Knowledge of Consciousness

The task now, in the broader context of the thesis, is to discover in what way the phenomenological reduction is related to our ongoing enquiry into the state of pure, pre-reflective, thought-free consciousness. To begin, I must again consciously practice a negation of the entire stream of reflective thinking that normally occupies my "mental space" as soon as my attention is disengaged from the things and events occurring in the world. I return to a state of pure receptivity. Whatever happens, happens, without my judging, commenting, valuing, remembering, associating, and so on. There is an alert, enduring, inner stillness. I find I can remain in this state, and that I can emerge again into an attentive reflection on the *trace* of this state as it is retained in the stream of my past experience. On the basis of this reflective retention, I can form an "image" of what it is to be without thought. But I do not mistake the image for the reality – for the reality is immediately present as soon as my reflection subsides.

The question now arises as to whether, in negating thought, I am inadvertently practicing a transcendental epochē and suspending my acceptance of the actuality of the world. I find I can hold this question, as a pure meaning, and then *see* how it stands with my immediate, thought-free consciousness 'now.' Initially, I find I have no knowledge of my accepting or not accepting the actuality of the world as something transcendent of my immediate experience. I do not know where to look in order to find something "transcendent." I cannot even see that I am having "an

experience." There is the table in front of me. If you were to ask if it were an "actual" table, as opposed to, say, an imaginary table, I can immediately see that it is an actual table. If you were to ask if there is something more to the table than its actually appearing in front of me, in the sense of its having an independent existence that is causing its appearance, then I would say that I cannot see an answer to that question. I am not thinking, I am simply looking into my consciousness-of-the-world. If you were to ask me to look into the how of the manner of givenness of my consciousness-of-the-world, for example, how the table is given to me in adumbrations as I move around it, how I expect to see the leg of the table revealed as I move my head to one side, I would say, yes, I see the manners of givenness of the table progressing flowingly and harmoniously.

However, the deeper question is whether the pre-reflective consciousness that opens up and enables my consciousness to be a consciousness-of-the-world is in some way transcendent of the world as it appears for me. Here I *am* able to answer (on the basis of a *direct knowledge*): my pre-reflective, thought-free consciousness, that consciousness that encompasses my consciousness-of there being a table in front of me, *that* consciousness is simply conscious. *It* is uninvolved with the contents that are passing through it. It *gives* what is experienced without being a part of that experience (it rather *embraces* the experience). It is not personal. It *gives* me, insofar as I appear as a body, with its sensations, the feeling of sitting on the chair. *It* could be described as an uninvolved, disinterested spectator (*if* I were to reflect on or objectify it). As it is, without reflection, it is an immediate knowledge. Knowledge of what? Of itself, of *now* – not just of the now of the present moment of the world, but the now of the present moment of consciousness, that itself includes the temporal horizons of what is passing and the expectation of what is to come.

If we examine the language of the phenomenological reduction, it is clear that Husserl and Fink have both encountered a transcendental state that is uninvolved with the intentionality of my consciousness-of-the-world. It is this state that has imbued them with the sense of having broken free from what we would call the normal, everyday state of consciousness. Husserl now emphasises that he stands *"above* the world," and has attained "a complete transformation of all life" comparable to "a religious conversion." These phrases all indicate that a *change in the centre of consciousness* has occurred, such as we achieve in the negation of thought, and not just a change in reflective attitude.

However, Husserl does not speak in terms of negating thought. In fact, the phenomenological reduction is supposed to elevate the meditating philosopher into a state of transcendental *reflection,* producing what Fink calls the transcendental onlooker or the transcendental witness:

an entirely new state of reflection with its own 'I' to be distinguished from the 'I' of the human individual in the world and the 'I' of the constituting transcendental ego – although all three I's are understood to possess an ultimate unity. This transcendental onlooker is a *thinking, reflecting* onlooker, albeit that it thinks transcendentally, i.e. *within* the phenomenological reduction. The aim of the phenomenological reduction is therefore not our aim. Husserl seeks after a state of pure *reflection* and uses the epochē to "forbid" the encroachment of a worldly understanding. In contrast, we seek to annul reflection itself.

Nevertheless, the phenomenological reduction must hover in the *vicinity* of the state pure pre-reflective, thought-free consciousness. It is the presence of this state that annuls the "natural attitude" and reveals the phenomenon of the world (the things themselves) that Husserl wishes to reflectively investigate.

Appendices

Appendix A

An Essential Difference

Wheeler and Heidegger on the Relationship Between Science and Philosophy

[This is an extended version of a paper originally presented at *Reconstructing the Cognitive World: A workshop with Michael Wheeler,* Goethe University Frankfurt am Main, 2010.]

A.1 Reconstructing the Cognitive World

Michael Wheeler, in his book *Reconstructing the Cognitive World,* analyses the development of embedded-embodied cognitive science in the light of underlying and largely unacknowledged philosophical differences about the constitution of human agency. On one side he sees orthodox computational cognitive science, despite its non-dualist physicalist credentials, as holding to Cartesian conceptions of an abstract, disembodied reason deliberating over decontextualised representations of the world. On the other side, he sees modern-day embodied-embedded cognitive scientists going beyond such Cartesianism and embracing concepts of human agency that have more in common with Heidegger's account of Dasein in *Being and Time.* By bringing to light and criticising the Cartesian assumptions of the computationalists and by pointing out and clarifying the connections between embodied-embedded thinking and Heideggerian existential phenomenology, Wheeler aims to lay the "foundations of a genuinely non-Cartesian cognitive science" (Wheeler, 2005, p. 16).

In the process, Heidegger is presented as a realist who holds that modern science provides genuinely objective epistemic access to independently *Real* entities. On this basis, Wheeler argues that Heidegger would have no objection to the incorporation of his account of Dasein in the

broad framework of contemporary cognitive-scientific explanation. According to Wheeler, such explanation is:

> . . . a species of empirical explanation in which the ultimate goal is to map out the sub-agental elements (e.g., the neural states and mechanisms, or the functionally identified psychological subsystems) whose organization, operation, and interaction make it intelligible to us how it is that unmysterious causal processes (such as those realized in brains) can give rise to psychological phenomena that are genuinely constitutive of agency and cognition (Wheeler, 2005, p. 127).

Within this framework, cognitive scientists necessarily make assumptions about *what* the relevant psychological phenomena are and *how* they are constitutive of agency and cognition. It is in these assumptions that Wheeler sees the hidden hand of Descartes and the need for a thorough-going philosophical clarification. By settling on this task, Wheeler explicitly accepts the validity of the cognitive-scientific framework and limits his clarification to operate *within* its confines. In the context of such an enquiry, it is clear that Heidegger's concepts of the *ready-to-hand,* the *unready-to-hand* and the *present-at-hand* provide valuable insights into how an embodied-embedded intelligence can cope with the world, and that these insights are suggestive of the kinds of neural mechanisms and functional subsystems that could underpin such coping behaviours.

Nevertheless, Wheeler's approach supposes that Heidegger's philosophy can be incorporated into a contemporary cognitive science framework without itself being fundamentally transformed. To support this supposition, Wheeler looks at a number of passages from *Being and Time* where Heidegger appears to endorse a view that philosophical analysis can be usefully employed to clarify (purify) the constitutive assumptions that already prevail within a particular science. Taken together with the evidence for a realist reading of Heidegger, Wheeler concludes that his philosophical clarification remains basically Heideggerian in outlook, and that those differences that do arise (for example, in relation to the animal's possession of world) are minor adjustments rather than fundamental divergences.

In this paper we shall argue that there *are* important differences between Heidegger and Wheeler about the proper relationship between science and philosophy. To demonstrate this, we trace the lineage of Wheeler's thought back to the philosophy of mind of Daniel Dennett. Here it becomes clear that the project of incorporating Heidegger into cognitive science is part of a larger project that aims to reconcile science with human self-conceptions, and to do so in

scientific terms. It is *this* philosophical perspective that determines the framework and methodology of Wheeler's cognitive science project, and it is here that more obvious differences between Wheeler and Heidegger begin to emerge. On this basis we revisit several of the passages that Wheeler uses to support his view of a Heideggerian philosophy-science nexus and subject them to further criticism. Our analysis will draw particularly on Heidegger's discussion of biological science in the *Fundamental Concepts of Metaphysics* where he shows more clearly the kinds of transformations a Heideggerian perspective could effect on the theoretical constructs of a science (like cognitive science) that deals with living beings rather than the present-at-hand objects of physics.

A.2 The Non-Reduction of Psychological Phenomena

In the first chapter of *Reconstructing the Cognitive World,* Wheeler declares a basic commitment to naturalism, which he defines as the position:

> (i) that physicalism is true, and (ii) that philosophy is continuous with natural science. [...] In my book, physicalism amounts to the ontological claim that there is ultimately nothing but physical stuff. It does not impose the additional explanatory condition that every worldly phenomenon be ultimately explicable by physical laws. [...] I read continuity with natural science in the weakest possible way, that is, as mere *consistency with* natural science, a reading that makes room, in principle, for multiple modes of explanation. Thus the view I advocate does not demand reductionist explanations of psychological phenomena' (Wheeler, 2005, p. 5).

This view is identified with the position outlined in Elton's book on Daniel Dennett (Elton, 2003), with Wheeler adding that "if there is a clash between philosophy and some *final* natural science, then it is philosophy that should give way" (Wheeler, 2005, p. 6). The briefness of Wheeler's statement on his naturalistic position indicates that he does not intend to examine the legitimacy of this foundation but rather to examine Heidegger in a *naturalistic light.* However, for our purposes, the legitimacy of this foundation *is* pertinent and we therefore need to examine Dennett in more detail.

In the passage previously quoted, Wheeler is clearly granting *some* form of reality to psychological phenomena, in that he rejects reductionist explanations, while at the same time holding

that "there is ultimately nothing but physical stuff." The question here is exactly *what* kind of reality we can grant to psychological phenomena when everything is ultimately physical. If we follow Wheeler's directive to Dennett, this issue is transformed into the question of what it means for something to be *reduced* to something else. Here, a reduction is considered to be a lawful translation of terms between two theories, such that terms in the first (higher) theory can be expressed via suitable bridge laws as equivalent to terms in the second (lower) theory. So, to take a typical example, we can lawfully reduce the concept of temperature in a gas down to the movement of the molecules that make up the gas. In Dennett's philosophy, psychological states and processes are represented as intentional phenomena, which are revealed to a third-person perspective by taking up an *intentional stance* (Dennett, 1987). By looking at certain complex systems, such as animals and humans, from an intentional stance, we can understand their behaviour in terms of intentional concepts, for example, as an agency possessing beliefs and desires and acting for reasons. Dennett's non-reductionism consists in the view that these intentional phenomena cannot be formally reduced via bridge laws to specific entities that exist at a physical or functional level. This failure of reduction is to be partly explained by the fact that the same intentional phenomenon can be realised in many different nervous systems and mechanisms, and partly by the observation that even within particular systems there need not be a direct relationship between the kinds of entities found at a physical or functional level and the kinds of behaviours manifested at the level of intentionality.[1] What matters for Dennett is the *behaviour,* for it is the behaviour that elicits the intentional interpretation and not any inner organisation of parts and functions. So, while inner organisation can explain the precise behaviour in a *particular* instance of observed intentionality, there is no lawful connection *in general* that can specify just how a belief, for example, can be defined in terms of lower level physical or functional structures.

One way to deal with this failure of reduction is to deny any reality to intentional phenomena (as in the eliminativism of Churchland (1988)). Another is to hold that there are bridge laws, but that a suitable translation has yet to be worked out that can handle all possible realisations (as in the representationalism of Fodor (1975)). However, Dennett holds that there *are* reasons for certain behaviours, that these reasons are new, objective phenomena that have arisen out of the evolution of the species, and that without such reasons we cannot adequately explain or predict the behaviour of complex intentional systems. He therefore grants a kind of relative reality to

[1] Elton illustrates the lack of connection between lower level entities and higher level behaviours with the example of a "hidden hand" guiding a market to equilibrium (Elton, 2003, p. 88).

intentional phenomena, saying that they are "real enough" (Elton, 2003, p. 92), while at the same time acknowledging that any particular intentional state exists as a pattern that is realised in the causal physical system that underlies it.

Dennett's position on this relative reality of intentional states is best understood as arising from a definite project, which Elton describes as the reconciliation of science and our self-conception. So Dennett is not engaged in *proving* that intentional states possess some form of reality, rather he is showing *how we must think* about intentional states if we are to retain them as real while also accepting the metaphysical outlook of modern science.

A.3 The Cognitive Science Framework

The parallels between Wheeler and Dennett extend to their conception of the framework of cognitive-scientific explanation. Here Wheeler divides the domain of enquiry into agential and subagential levels. This bears close resemblance to Dennett's distinction between the personal and the subpersonal, and his development of the physical, design and intentional stances. The difference is that Wheeler, following Elton, expands Dennett's conception of the personal to form an agential/subagential distinction, where agency includes not just human self-reflective reasoning but any system that can reasonably be interpreted as possessing agency from the intentional stance. The subagential level then refers to how these systems appear from a physical or design (functional) stance.

Within this framework, behaviours appearing at the agential level that can be explained in terms of intentional concepts (e.g. by reasoning about psychological states), can be connected to causal explanations involving subagential mechanisms. The point is that such explanation does not need to formally reduce the intentional/psychological account of agency to the subagential level. Rather, by *flipping* between the two ways of viewing a system, the aim is to show how the subagential causal mechanisms *unmysteriously* enable agential level behaviours to appear. This flipping is achieved by blurring the agential/sub-agential line and allowing that subagential subsystems can themselves be viewed as possessing intentionality (or meaning or content) within the context of their being a part of a larger agential system. However, the precise way that subagential subsystems can be understood as possessing content remains (perhaps deliberately) unclear in both Dennett and Wheeler. Elton interprets Dennett as simply attaching intentional labels to subagential components in a way that should not be taken literally, whereas Wheeler thinks an en-

tirely metaphorical understanding of subagential content "seems unnecessarily strong" (Wheeler, 2005, p. 301).

Finally, both Wheeler and Dennett take it as given that, as an empirical science, cognitive-scientific enquiry requires the taking up of a third-person perspective in relation to the phenomena to be studied. Dennett makes this explicit in his characterisation of heterophenomenology in *Consciousness Explained* (1991, p. 98). Here, first-person accounts of phenomenological experience are to be taken as narratives concerning fictional entities that can only be considered real on the basis of corroborating third-person evidence (such as the identification of an underlying physical referent in the brain). Wheeler similarly agrees that the final court of appeal for his Heideggerian cognitive science is the third-person arena of empirical results, where it is "the concrete empirical success of a cognitive science with empirical credentials, that, if sustained and deepened, would ultimately vindicate a Heideggerian position in cognitive theory"(Wheeler, 2005, p. 189).

A.4 Introducing Heidegger

In *Reconstructing the Cognitive World* Wheeler's aim is to introduce Heidegger's philosophy into the overarching framework of contemporary cognitive science in such a way that they both remain *essentially* unchanged but mutually enriched. This is achieved by reading Heidegger as agreeing that the proper task of philosophy in relation to an empirical science (such as cognitive science) is to provide *constitutive* explanations of the target phenomena (in this case the phenomenon is human agency, as conceptualised within Wheeler's cognitive science framework). Wheeler supports his interpretation via an appeal to Heidegger's discussion of anthropology, psychology and biology in *Being and Time*:

> In suggesting that anthropology, psychology, and biology all fail to give an unequivocal and ontologically adequate answer to the question about the *kind of Being* which belongs to those entities which we ourselves are, we are not passing judgment on the positive work of these disciplines. We must always bear in mind, however, that these ontological foundations can never be disclosed by subsequent hypotheses derived from empirical material, but that they are always 'there' already, even when that material simply gets *collected.* If positive research fails to see these foundations and holds them to be self-evident, this by no means proves that they are not basic or that they are not problematic in a more radical sense than any thesis of positive science

> can ever be (Heidegger, 1962/2008, p. 75).

Here Heidegger indicates that these empirical (positive) sciences have not arrived at an "ontologically adequate answer" as to the "*kind of Being*" that they are studying, and that they will be unable to uncover their own radically problematical ontological foundations on the basis of empirical work alone. The first task for philosophy is therefore to reveal those ontological foundations that positive research is unable to see. For Wheeler, in the context of contemporary cognitive science, this involves explicating and criticising the Cartesian constitutive assumptions that lie behind orthodox cognitive science's computationalist conception of human agency.

Philosophy's second task is then to discover an "ontologically adequate answer to the question about the *kind of Being* which belongs to those entities which we ourselves are." Here Wheeler takes it that just such an answer is provided in the interpretation of Dasein from *Being and Time.* The third task is then to develop, or discover, on the basis of this interpretation, the most appropriate ontological foundations upon which the future empirical work of the positive science in question can be based. In a significant move, Wheeler takes it that current embedded-embodied cognitive science, as a result of theoretical reflection and empirical discovery, has *already* discovered these Heideggerian ontological foundations (albeit in an unclarified or impure form). Consequently, Wheeler's attention becomes focussed on the philosophical clarification of these assumptions and the demonstration of their connection back to Heidegger's original interpretation of the structure of Dasein.

On this interpretation of the role of philosophy in relation to positive science, it is important to see that Wheeler takes the enquiry into the ontological foundations of cognitive science to be *no more* than an investigation into what constitutes human agency. What remains unquestioned is the ontological framework within which the concept of human agency is first developed, that is, the physicalist, third-person perspective, and the agential/subagential distinction from which the understanding of mind and cognition as agency emerges. Wheeler goes so far as to credit Heidegger with having divided the domains of science and philosophy in this way:

> ... Heidegger's approach is to disentangle two intellectual challenges that, in the context of the study of mind, emerge as (i) the identification and clarification of the constitutive character of human agency (in Heideggerian terminology, the Being of human agents), and (ii) the empirical investigation of how human agents (and their collective social groups) work causally so as to realize that character. These two chal-

> lenges correspond naturally to two different modes of explanation, that we can call the *constitutive* and the *empirical.* For Heidegger, it often seems that constitutive explanations are distinctively the business of philosophy – in particular, of a disciplined and systematic *phenomenology* – whereas empirical explanations are distinctively the business of science. Moreover, Heidegger appears to hold this division as a matter of principle. Thus he writes that "ontological foundations [of a science] can never be disclosed by subsequent hypotheses derived from empirical material . . . they are always 'there' already, even when that empirical material simply gets *collected*" (p. 75, original emphasis). What this tells us is that, for Heidegger, constitutive explanations cannot be reduced to empirical explanations (Wheeler, 2005, pp. 125–126).

What Wheeler is assuming is that Heidegger would equate an enquiry into the ontological foundations of cognitive science with "the identification and clarification of the constitutive character of human agency." The problem is that Wheeler *already* has a pre-understanding of the *being* of human agency formed within the framework of cognitive science – that is, within the framework of a scientific understanding. He therefore understands the task of philosophy as providing an ontological inventory of the kinds of things that make up such a scientifically conceptualised human agency. Whereas Heidegger's task in *Being and Time* was to investigate the being of Dasein (human agency) from a radically original perspective that *precedes* any such scientific conceptualisation. This is made clear in the introduction discussion of the question of being in *Being and Time*:

> The question of Being aims therefore at ascertaining the *a priori* conditions not only for the possibility of the sciences which examine entities as entities of such and such a type, and, in so doing, already operate with an understanding of Being, but also for the possibility of those ontologies themselves which are prior to the ontical sciences and which provide their foundations. *Basically, all ontology, no matter how rich and firmly compacted a system of categories it has at its disposal, remains blind and perverted from its innermost aim, if it has not adequately clarified the meaning of Being, and conceived this clarification as its fundamental task* (Heidegger, 1962/2008, p. 31).

A.5 Heidegger's Realism

One response to the complaint that Wheeler fails to question the deeper ontological foundations of cognitive science is to argue that Heidegger's investigations in *Being and Time* actually support a scientific conception of human agency, at least within the context of cognitive science, so further examination is unnecessary. Wheeler implicitly takes this line in his interpretation of Heidegger as a scientific realist. The interpretation begins with Heidegger's distinction between *the Real* and *Reality* in *Being and Time.* Here 'the Real' refers to entities as they 'are' in themselves, that is, independent of being encountered by Dasein. The terminology is difficult because Heidegger only allows that there is being insofar as Dasein, as understanding of being, exists. So, without Dasein, we cannot properly speak of something being, even of it being 'in-itself.' Nevertheless, Heidegger does not want to say that there is nothing beyond Dasein and its understanding of being. So, beyond that, 'is' the Real.

The basic issue for Wheeler is whether Heidegger's acknowledgement of the Real as independent of Dasein can be used to justify his conception of the framework of cognitive-scientific explanation. Firstly he argues that science gives us epistemic access to "fully objective, agent-independent entities and properties" (Wheeler, 2005, p. 152), but is careful to qualify that what is revealed are only "mathematically describable causal properties" (Wheeler, 2005, p. 153) of present-at-hand objects. Wheeler goes on to criticise Dreyfus's interpretation of Heidegger's scientific pluralism, which Wheeler takes as denying a special status to modern science because it allows that "two scientific theories that contradict each other might conceivably be equally valid ways of understanding nature" (Wheeler, 2005, p. 155). However, a closer reading of Dreyfus shows that Wheeler has oversimplified the issue. Importantly, Dreyfus's Heidegger is only proposing that there are different ways that science can conceptualise nature. Firstly, Dreyfus points out that, according to Heidegger, the 'objectness' of science is only one way of conceptualising nature:

> What is represented by physics is indeed nature itself, but undeniably it is only nature as the object-area, whose objectness is first defined and determined through the refining that is characteristic of physics and is expressly set forth in that refining. Nature, in its objectness for modern physical science, is only *one* way in which what presences – which from old has been named as *physis* – reveals itself (Heidegger, 1977b, pp. 173–174).

Dreyfus then finds Heidegger asserting that there are many possible forms of scientific projection, none of which can be considered as more correct than another:

> [We cannot] say that the Galilean doctrine of freely falling bodies is true and that Aristotle's teaching, that light bodies strive upward, is false; for the Greek understanding of the essence of body and place and of the relation between the two rests upon a different interpretation of entities and hence conditions a correspondingly different seeing and questioning of natural events. No one would presume to maintain that Shakespeare's poetry is more advanced than that of Aeschylus. It is still more impossible to say that the modern understanding of whatever is, is more correct than that of the Greeks (Heidegger, 1977a, p. 117).

So Dreyfus is *not* proposing that two contradictory scientific theories could both be correct. He is rather attributing to Heidegger the view that two scientific theories that conceptualise nature in different ways can end up with different explanations of natural phenomena. This does not mean that two such theories need contradict each other, once the underlying conceptions are taken into consideration. Wheeler criticises this view by claiming the entailment that "alchemical chemistry and modern chemistry might both be true" (Wheeler, 2005, p. 155), with the unspoken assumption that such a possibility must be false. However, if we seriously envisage a present-day (rather than medieval) alchemy, then such a discipline, in whatever form it took, would have to *take into account* the findings of modern chemistry, and would therefore be careful to point out that its theories refer to elements that are not conceived in a purely physicalist projection. Hence there *need* be no contradiction at the level of theory. Rather, the contradiction would lie at the level of ontological assumption. Wheeler's final position is to say that accepting such alchemical ontological assumptions is "counterintuitive." However, there is an underlying circularity here, in that being counterintuitive is finally equated with not accepting the ontological primacy of the third-person, conceptualisation of modern science.

Wheeler's difference with Dreyfus, and ultimately with Heidegger, over the status of modern science is crucial for our present purposes, for if there is no privileged form of scientific conceptualisation, then Wheeler's conceptualisation of human agency itself becomes questionable. That is, it becomes just one of many possible conceptualisations and not necessarily the most suitable.

A.6 The Conceptualisation of Human Agency

Until now we have not fully explicated the ontological foundations of Wheeler's conceptualisation of human agency. In Heidegger's terms, we could say this conceptualisation has been guided by a certain *care.* Initially, this care is declared quite openly: it is that "physicalism is true" where "physicalism amounts to the ontological claim that there is ultimately nothing but physical stuff" and that "philosophy is continuous with natural science" (Wheeler, 2005, p. 5). The ontological commitment here is to nature as it is revealed through modern physical science. This revealing is what shows up when we assume the third-person observational stance, which limits us to what can be reliably confirmed in an external, objectified way, that is, to what is *measurable* (quantifiable). The physical is then defined in terms of entities that are capable of producing *measurable effects* and physicalism is the position that only what is capable of producing such effects is ultimately real. This scientific idea of the measurable is bound up with a search for lawfulness, where what is measured is understood in terms of an underlying conception of material cause and effect. It is here that Wheeler understands science as revealing the "mathematically describable causal properties of entities" and he explicitly identifies the Real with this "underlying causal structure" (Wheeler, 2005, p. 153).

In taking issue with Dreyfus on the question of there being other equally valid modes of scientific conceptualisation, Wheeler sees modern physical science as having discovered an ultimately correct *scientific* conceptualisation of the Real. This ultimate status is not argued for directly, but is supported by references to the wide acceptance of physicalism amongst other contemporary philosophers and researchers in cognitive science, and by the conviction that the alternatives of dualism and idealism are not worth taking seriously. Nevertheless, Wheeler is not arguing that physical science gives us an *ultimate access* to the Real, rather that we have a partial access to the causal structure of the Real.

The question is, how does this privileging of the scientific, third-person perspective determine Wheeler's conceptualisation of human agency? Firstly, we can say that Wheeler's motivation (his care) is to found his conceptualisation on the authority of physical science. This authority is taken to entail the acceptance of some form of physicalism. Thus, the correctness of the approach of physical science is assumed beforehand, and acts as the foundation for all subsequent enquiry. The task is then to explain mind and cognition in terms of this physical foundation, without introducing anything that cannot also be understood as ultimately physical. The care is to make

such a conception of what is real *intelligible* without recourse to the mysterious – where the mysterious is anything that cannot be ultimately understood as deriving from the physical. In other words, the task is the development of a set of concepts that make it possible to think about mind and cognition as founded on an ultimate physicalism without falling into obvious contradiction, while also making it intelligible that nothing fundamental has been left unexplained.

The strategy for such an explanation is to divide the region of interest into agential and subagential categories and to make the subagential primary by explaining the agential in terms of the subagential. The category of the subagential already follows the scientific perspective, in that the physical-as-scientifically-measurable falls on this side, while those aspects of Reality (as understood by Heidegger) that this physical conceptualisation does not encompass are placed on the agential side. This means that being, as the being of human experience, is immediately placed in an ontologically subordinate position, something to be explained in terms of physical cause and effect, something that is *enabled* by the physical.[2] Within this framework, the question of being, as it was developed by Heidegger, belongs in the region of the agential, and can no longer question the overall conception of Wheeler's physicalism. Instead, the scope of Heidegger's question has to do with mapping out the constitution of human agency.

In this way Wheeler *reverses* the primacy of Dasein as the ground and source of the understanding of being. Instead, he substitutes a scientific understanding of being as ultimately physical, and places Heidegger's enquiry into being within this overarching physicalism. Now, much hinges on what 'ultimately' physical might mean in this context. As we have already discussed, Wheeler's physicalism allows for the non-reduction of psychological phenomena and so grants a kind of relative reality to agential level phenomena. Nevertheless, there is still an assertion that physicality is foundational. One interpretation is that Wheeler is *not* asserting the primacy of the being of the physical (as the present-at-hand) but is trying to indicate, within the confines of language, that the *Real* is ultimately physical. On this basis, the subagential has to do with the Real cause and effect structure of entities, whereas the agential has to do with the various modes of Reality of agential experience. So, only at the agential level do we encounter being as the being-in-the-world of Dasein, of the present-at-hand, the ready-to-hand, and so on. Once the third-person perspective of science is assumed, we emerge into the Real, in such a way that

[2]The idea that the subagential *enables* agential level phenomena is linked to John McDowell's idea of enabling explanation (McDowell, 1994a). Wheeler (in a footnote pp. 300–301) shows some sympathy with McDowell's position but does not clarify where he stands on McDowell's much more Heideggerian understanding of perceptual experience as a direct contact with the world.

we can consider Dasein (human agency) from the 'outside.' There are several passages from *Reconstructing the Cognitive World* that support this view. For example, when dealing with Varela, Thompson and Rosch's contention that scientific enquiry "must itself be a product of the structure of [the scientist's] own cognitive system" (Varela, Thompson, & Rosch, 1993, p. 11), Wheeler responds that "subagential explanations are generated from the standpoint of the detached attitude, and thus emerge as fully realist in character" (Wheeler, 2005, p. 157). In another passage, Wheeler connects the being of the present-at-hand with the Real, making it clear that he takes scientific realism to be describing entities as they are in themselves and not from any observer relative perspective:

> Scientific practices paradigmatically reveal the mathematically describable causal properties of entities, properties that, as the present-at-hand, are precisely not related to any particular network of everyday significance. [...] Thus the Real (entities, or, more accurately, the underlying causal structure of entities) are independent of everyday significance; but the fact that the Real are *intelligible* as being independent of everyday significance requires there to be the phenomenon of intelligibility (Reality). In other words it requires there to be a human agent (Wheeler, 2005, p. 153).

However, the issue here is not whether scientific practices have attained to an observer independent perspective in relation to describing the causal properties of entities. The issue is whether taking such a perspective is appropriate when the object under consideration is human agency (Dasein). For it can easily be agreed that the third-person perspective is paradigmatically suited to the revealing properties of *physical* objects, that is, objects with the mode of being of the present-at-hand. Wheeler's assumption is that such a perspective is also appropriate for explaining (revealing) entities with the mode of being of Dasein, that by treating Dasein (at the subagential level) as if it were a present-at-hand object, we can help to reveal it as it is in itself, rather than, for instance, revealing it as something it is *not.* Of course, Wheeler is *at the same time* considering Dasein from an agential level that does allow the various modes of being of Dasein to emerge. But the issue is not that the being of Dasein is ignored, it is that the third-person, present-at-hand view has been made *foundational,* that in terms of which all else is to be explained. Finally, Wheeler's contention would have to be that insofar as cognitive science is a science then it must take up a scientific (detached) perspective towards its subject matter. But here we return full circle, for this is Wheeler's care, to introduce a Heideggerian cognitive science that leaves the detached attitude

unchallenged. Whereas our care is to question how far such an attitude can be reconciled with Heidegger's radical questioning of being.

In addition, even though Wheeler admits Heidegger's analytic of Dasein as relevant at the agential level, the third-person perspective still determines how that analytic is understood. So, to start with, the very term Dasein is reinterpreted from an external viewpoint as a concept with a certain extension:

> The word "Dasein" is a technical term used by Heidegger to pick out those creatures who enjoy a certain specific and distinctive form of existence. On planet earth at least, that form of existence is realized only by human beings. [...] Having drawn attention to the unfamiliar term "Dasein," I shall now seek to to avoid the sense of mysticism that it can easily engender by immediately dropping it from my exposition altogether ... and by speaking instead simply of "human agency" and "the human agent" (Wheeler, 2005, pp. 121–122).

In Wheeler's third-person conceptualisation, Dasein is immediately understood *behaviourally,* that is, as an agent that is the locus of externally observable behaviour, rather than as a *being* whose very existence is its 'to-be' (Heidegger, 1962/2008, p. 67). Similarly, being itself becomes externally understood as a capacity or an ability of the human agent to make the world intelligible:

> "Being" does yeoman service for Heidegger, indicating both (i) intelligibility (the property of making sense or of being meaningful), and (ii) the background understanding on the basis of which entities may show up as intelligible (meaningful) (Wheeler, 2005, p. 124).

The care of this conceptualisation of human agency is to encircle Dasein within a third-person perspective that *places* it in an agential and subagential framework of understanding. Once encircled, the being of Dasein is understood in terms of behaviours, and particularly in terms of phenomenological narratives describing 'first-person' investigations of Dasein. From this perspective we can understand Heidegger's philosophy as just such a narrative, one that indicates certain structures of experience (modes of being) that can act as clues to developing a better understanding of the subagential level mechanisms and subsystems. In a complementary way, empirical validation of subagential models can provide evidence of the correctness of those structures described in the phenomenological narrative.

The task here is not to argue that Wheeler's interpretations of Dasein and being are incorrect, but to indicate that the third-person perspective inheres in the language of human agency in such a way that Dasein becomes *invisible.* This language makes Dasein into an object, and consequently makes oneself (again Dasein) assume the role of a necessarily implied subject, standing-over-against Dasein-as-object. This standing-over-against is possible in relation to present-at-hand objects and arises out of Dasein naturally, once it looks at things theoretically. But Dasein is the being that adopts this mode of being towards the Real that reveals the present-at-hand. And Dasein is not a present-at-hand object in the first place; its being lies in its 'to-be.' Once Dasein adopts a third-person perspective towards understanding *itself* there is a circularity involved, in that Dasein exists as Dasein while at the same time envisaging that it stands outside Dasein. This standing-outside can only have an 'as-if' status, for Dasein remains as Dasein throughout. An awareness of this circularity is important, for Dasein only shows up as Dasein for Dasein, and not for any third-person perspective (which necessarily sees Dasein 'from the outside'). Put another way, once Dasein becomes objectified as human agency, then its being becomes concealed. The care to see Dasein *exclusively* in third-person terms forgets that Dasein is not a present-at-hand object, but is the source and origin of all perspectives, even one that seeks to become observer independent. Such independence is possible in relation to an object that is *not* Dasein, but once Dasein is the object of enquiry, then the observation itself can no longer be taken for granted: this too must be clarified.[3]

A.7 The Incomprehensibility of Physicalism

As we have already discussed, Wheeler's physicalism can be taken in two ways: (i) that the present-at-hand *as a mode of being* is ultimate or (ii) that the Real itself should be understood as ultimately "made of physical stuff." Of these, we can take it that Heidegger did not mean to privilege the being of the present-at-hand over, for example, the being of Dasein. So we can reject

[3]It is worth noting how the idea of a third-person perspective creates the *illusion* of an opposed first-person perspective. This occurs when the third-person perspective believes that it *really* stands outside of Dasein, forgetting that Dasein itself has assumed the perspective in the first place. It then sees all other perspectives that Dasein assumes as occurring 'within' Dasein, as non-third-personal, or as first-personal. But as all perspectives are revealings of being to and of Dasein, it is rather the case that *all* perspectives that Dasein assumes are first-personal in this sense, so the distinction itself serves no purpose. The real distinction, as Heidegger sees it, is not between the third person and the first person, but between there being a person or not being a person at all (in the sense of a subject in a subject/object relationship).

(i). In relation to (ii), the position that the Real is ultimately physical, this too is hard to support. For Heidegger's conception of being is that being cannot 'be' without Dasein, and hence that we cannot sensibly speak of the being of the Real without implicating Dasein as the place where such being is revealed. For example, he says: "Of course only as long as Dasein *is* (that is, only as long as an understanding of Being is ontically possible), 'is there' Being. When Dasein does not exist, 'independence' 'is' not either, nor 'is' the 'in-itself.' " (Heidegger, 1962/2008, p. 255).

The most likely interpretation is that Wheeler means to think of the Real as ultimately physical in some way that does not implicate being. He indicates as much in his use of the term 'occurs' in relation to the idea of pre-history (before human agency appeared on earth):

> ... the different ways in which entities make sense to us are dependent on the fact that we are human agents, creatures with a particular mode of existence; but it does not follow from this that those entities require the existence of the human agent in order just to occur (in an ordinary, straightforward sense of "occur"), only that they require the human agent in order to be *intelligible* as entities that just occur (Wheeler, 2005, p. 154).

One can sense here Wheeler's impatience to assert on the basis of common sense, that, of course there were things, ordinary physical things 'occurring' when there was no one there to see them. However, the reference to 'occur' (in order to avoid speaking of being) does not get Wheeler off the hook. For what ordinary, straightforward sense of occur is there that does not involve the being of the thing that occurs? Wheeler's assertion simply leads back to his original ontological position, that the Real is ultimately physical and hence that there is an ultimate objective spacetime in which physical entities ultimately 'occur.' Whereas, for Heidegger, the Real is what lies 'behind' being, a way of referring to what cannot be envisaged, because all envisaging is a conferral of being. We cannot make the Real intelligible except through an imagining of a space and time that exists independently of Dasein, and we have no warrant to assume the existence or the nonexistence of such a dimensionality. It *could* be that the universe exists as a completed piece of mathematics in a universal mind that only gets spatio-temporal being conferred on it through the existence of Dasein as being-in-the-world. Or it *could* be that ultimately there is only physical stuff in an objective spacetime. Heidegger's point is that our understanding is finite; it is an understanding of *being* that cannot go *beyond* being. That does not mean that we do not have contact with the Real, but it does mean that that contact is in the *form* of being. To think of what

lies behind being is to try and think of something that has no form, that is, to think of nothing. Heidegger makes this clear in describing nature, once it becomes the unworlded present-at-hand of scientific investigation, as ultimately "incomprehensible":

> Every explanation, when we speak of an explanation of nature, is distinguished by its involvement with the *incomprehensible.* It can be flatly stated that *explanation is the expository interpretation of the incomprehensible,* not so that this exposition would let us comprehend the incomprehensible, for it remains incomprehensible in principle. *Nature is what is in principle explainable and to be explained* because it is in principle incomprehensible. It is *the incomprehensible pure and simple.* And it is the incomprehensible because it is the *"unworlded" world,* insofar as we take nature in this extreme sense of the entity as it is discovered in physics. This is connected with the fact that in this kind of explanation and discovery of the world as nature, nature is still investigated and interrogated only with regard to the presence of the entity in it; and this entity is admitted only insofar as it is determined by laws of motion which remain invariant, unaltered, always the same for every possible approach and regard under which the consideration of nature is placed. It should be observed here that all propositions and proofs given in physics or mathematics are certainly comprehensible as propositions, as discourse about something, but that about which they speak is itself the incomprehensible. As the incomprehensible, it is likewise the entity which simply does not have the character of Dasein at all, while Dasein is the entity which is comprehensible in principle. Since understanding belongs to its being as being-in-the-world, world is comprehensible to Dasein insofar as it is encountered in the character of meaningfulness' (Heidegger, 1992, pp. 217–218).

This passage is revealing of Heidegger's avoidance of any direct statement about how we are to 'ultimately' understand the Real and his related avoidance of labelling himself as either a realist or as an idealist. Such avoidance falls directly out of the understanding that being, although not a *creation* of Dasein, can only be revealed through the existence of Dasein. This allows Heidegger to say that *neither* realism or idealism are correct:

> Along with Dasein as Being-in-the-world, entities within-the-world have in each case already been disclosed. This existential-ontological assertion seems to accord with the thesis of *realism* that the external world is Really present-at-hand. In so far as

> this existential assertion does not deny that entities within-the-world are present-at-hand, it agrees – doxographically as it were – with the thesis of realism in its results. [. . .] But what distinguishes this assertion from realism altogether, is the fact that in realism there is a lack of ontological understanding. Indeed realism tries to explain Reality ontically by Real connections of interaction between things that are Real.
>
> As compared with realism, *idealism,* no matter how contrary and untenable it may be in its results, has an advantage in principle, provided that it does not misunderstand itself as 'psychological' idealism. If idealism emphasizes that Being and Reality are only 'in the consciousness,' this expresses an understanding of the fact that *Being cannot be explained through entities* [my emphasis] (Heidegger, 1962/2008, p. 251).

So, while Heidegger accepts a realism about the world as present-at-hand and allows that each entity has an 'in-itself' "independent of any apprehension of it" (Heidegger, 1992, p. 217), this does not warrant his acceptance of physicalism or the idea that all science must start from physicalist assumptions. Equally, this does not entail his rejection of Wheeler's conception of human agency, or the attempt to explain cognition in terms of physical cause and effect, *within the domain of cognitive science.* For Heidegger, the question is whether such a framework is the most *appropriate* conceptualisation – one that allows the subject matter of a science to "[show] itself in itself" (Heidegger, 1962/2008, p. 254). However, the important point (for now) is that Heidegger denies physicalism any ultimate or *a priori* validity, meaning that the grounding of Wheeler's cognitive science framework in the basic assumptions of modern physicalist science becomes *open to question*:

> The "fact" of the sciences, i.e., the factical subsistence of an understanding of being, which is necessarily contained in them as in all comportment towards beings, can neither be the authority that grounds their a priori, nor can it be the source for knowledge of that a priori. Rather, it can only be one possible occasion for pointing us toward the originary ontological constitution of, for example, history or nature. Such a pointer must itself remain subject to a constant critique that has already taken its guidelines from the fundamental problematic of all questioning concerning the being of beings (Heidegger, 1998b, pp. 104–105).

A.8 The Reversal

From this initial discussion, we have merely concluded that Heidegger would not necessarily have agreed with Wheeler on his mode of objectification of human agency. Certainly Heidegger agreed that it is the business of science to objectify nature and it is the business of philosophy to investigate the being of beings, and that philosophy can be of assistance to science in forming appropriate objectifications of nature (via an investigation of the being of the beings that comprise the subject area of the science in question). What has caused us to pause, is that the subject matter of cognitive science is not material nature, but mind and cognition and their relationship with the empirical investigation of the brain. As Dasein is implicated, we can no longer assume the appropriateness of a third-person, scientific objectification, that starts with the physical brain in a physical environment and moves from there to explain the being of mind and cognition.

Heidegger did not provide a straightforward indication of how to conceptualise this relationship between events occurring in the brain and the existence of Dasein as being-in-the-world. However, he did make it clear that science is *not* the standard or basis upon which the matter is to be decided:

> What happens here, that the tree stands there to face us, and we come to stand face-to-face with the tree? Where does this presentation take place, when we stand face-to-face before a tree in bloom? Does it by any chance take place in our heads? Of course, many things may take place in our brain when we stand on a meadow and have standing before us a blossoming tree in all its radiance and fragrance – when we perceive it. In fact we even have transforming and amplifying apparatus that can show the processes in our heads as brain currents, render them audible and retrace their course in curves. We can – of course! [...] But ... while science records the brain currents, what becomes of the tree in bloom? What becomes of the meadow? What becomes of the man – not the brain but of the man, who may die under our hands tomorrow and be lost to us, and who at one time came to our encounter? What becomes of the face-to-face, the meeting, the seeing, the forming of the idea, in which the tree presents itself and man comes to stand face-to-face with the tree?
>
> It will be said in rebuttal: What is the use of such questions concerning a state of affairs which everybody will in fairness admit immediately, since it is clear as day to all the world that we are standing on the earth and, in our example, face-to-face with

> a tree? But let us not slip too hastily into this admission, let us not accept and take this "clear as day" too lightly. For we shall forfeit everything before we know it, once the sciences of physics, physiology, psychology, not to forget scientific philosophy, display the panoply of their documents and proofs, to explain to us that what we see and accept is properly not a tree but in reality a void, thinly sprinkled with electric charges here and there that race hither and yon at enormous speeds. It will not do to admit, just for the scientifically unguarded moments, so to speak, that, naturally, we are standing face to face with a tree in bloom, only to affirm the very next moment as equally obvious that this view, naturally, typifies only the naïve, because pre-scientific, comprehension of things. For with that affirmation we have conceded something whose consequences we have hardly considered, and that is: that those sciences do in fact decide what of the tree in bloom may or may not be considered valid reality. Whence do the sciences – which necessarily are always in the dark about the origin of their own nature – derive the authority to pronounce such verdicts? Whence do the sciences derive the right to decide what man's place is, and to offer themselves as the standard that justifies such decisions? And they will do so just as soon as we tolerate, if only by our silence, that our standing face-to-face with the tree is no more than a pre-scientifically intended relation to something we still happen to call "tree." In truth, we are today rather inclined to favor a supposedly superior physical and physiological knowledge, and to drop the blossoming tree (Heidegger, 1968/2004, pp. 42–43).

This passage, while full of poetic resonance, also provides a basis from which to build a picture of what Heidegger would have to say about contemporary cognitive science. And that is, that we must start from the *encounter with the tree,* and not from our physical understanding of the brain. In effect, this *reverses* the approach that Wheeler has taken. A first response could be that Heidegger was being *philosophical* about his encounter with the tree, and that he certainly would not expect cognitive science to adopt such a directive. But, if we step back to Heidegger's treatment of the biological sciences in *The Fundamental Concepts of Metaphysics*, we find exactly the same reversal in his analysis of the *scientific* explanation of 'life':

> [What] the struggle within biology against physics and chemistry really means is that *"life" as such cannot in principle be grasped from within the perspective of these dis-*

> *ciplines.* Yet this also implies that we cannot start by explaining "living substance" in physico-chemical terms, only to find ourselves in the embarrassing position of having to admit some other factor later on when our calculations fail and we are left with an inexplicable residue. On the contrary, the delimitation of life must be accomplished on the basis of the fundamental character of living beings themselves as something that cannot be explained or grasped at all in physico-chemical terms. The task confronting biology as a science is to develop an entirely new projection of the objects of its enquiry. (Expressed from another point of view, which is not necessarily identical with what we have just said, the task today is to liberate ourselves from the mechanistic conception of life) (Heidegger, 1995, pp. 188–189).

Here Heidegger unambiguously stands against the idea of understanding "living substance" in mechanistic, "physico-chemical terms" because doing so leaves "an inexplicable residue." We can transfer this position directly to contemporary cognitive science, which similarly starts with a mechanistic, physico-chemical explanation of mind and attempts to explain the inexplicable residue (of consciousness and intentionality) in terms of the non-reduction of psychological phenomena. Heidegger makes the even stronger claim that the subject matter of biology must be seen in an entirely new projection, because the "fundamental character" of this subject matter "cannot be explained or grasped at all in physico-chemical terms." We should compare this with Wheeler's conception of cognitive-scientific explanation as:

> . . . a species of empirical explanation in which the ultimate goal is to map out the subagental elements (e.g., the neural states and mechanisms, or the functionally identified psychological subsystems) whose organisation, operation, and interaction make it intelligible to us how it is that unmysterious causal processes (such as those realised in brains) can give rise to psychological phenomena that are genuinely constitutive of agency and cognition (Wheeler, 2005, p. 127).

If Wheeler is to reconcile Heidegger's position on the biology of the 1920s with his own position on contemporary cognitive science, then further work is required. The most obvious course is to argue that Heidegger is standing against a form of reductionism in biology that does not implicate Wheeler's more sophisticated idea of the non-reduction of psychological phenomena. In other words, Heidegger is rejecting a form of biological eliminativism, analogous to Churchland's position in cognitive science, and may have looked more favourably on a conceptualisation of life

as a phenomenon that is *enabled* by mechanistic, physico-chemical processes within biological systems but that cannot be lawfully *reduced* to such processes. The question is whether such a conceptualisation would represent the kind of "entirely new projection" that Heidegger was proposing. In relation to biology, Heidegger's new projection centres on the "essential wholeness" of the living being as an *organism*:

> The fundamental thesis here is that everything that lives is an *organism.* [...] And this also implies that the concept of a "living substance," a vital mass or "life stuff," is a meaningless one. For the idea of "stuff" or "substance" in this sense specifically denies the character of the living being as an organism. [A living being's] organismic character is what determines the unity of this particular living being in each case. The unit of life is not the cell. The multicellular living being is not, as has been suggested, a community of cells. On the contrary, both unicellular and multicellular living beings alike possess a *unity* of their own in each case, that is, they have a specific *essential wholeness* by virtue of the fact that they are *organisms* (Heidegger, 1995, p. 212).

Initially, this conception of the organism appears analogous to Wheeler's conception of agency, in that both pick out the "essential unity" of a living being. However, Heidegger's organism starts to diverge from Wheeler's agency when Heidegger denies that an organism can be considered as a machine that additionally possesses "supra-mechanical functions":

> [Equipment] is what it is and in the way that it is only insofar as it is a *product* of human activity. And this implies that such production of equipment is only possible on the basis of what we have called *world-formation.* [...] If this is the case, then it is questionable whether we should attempt to grasp organisms as instruments or machines. And if this approach is excluded in principle, then it is also impossible to endorse that procedure in biology which begins by treating the living being as a machine and then goes on to introduce supra-mechanical functions as well. This procedure certainly does greater justice to the manifestations of life than any purely mechanistic theory. Yet it still misrepresents the central problem which we repeatedly forced to confront: that of grasping the original and central character proper to the living being (Heidegger, 1995, pp. 213–214).

Here Heidegger confronts and rejects a conceptualisation of the organism that maps closely with Wheeler's cognitive science framework. This biological conceptualisation starts with a mechanical explanation of the organism (corresponding to Wheeler's physical and functional understanding of the subagential level) and goes on to understand the organism in terms of supra-mechanical phenomena (corresponding with the unreduced psychological/intentional phenomena of the agential level). Heidegger rejects the conceptualisation of the organism as a machine on the basis that equipment is produced by human activity, whereas living organisms arise from out of a different essence. Of course, to a mechanistic conception, such a distinction does not matter, as it is easy to conceive of a machine that can assemble a replica of itself. If it could be agreed that such a machine is also a living organism, then the dispute would evaporate. However, as Heidegger's encounter with the tree in bloom already indicates, he does *not* think of living organisms as self-replicating machines:

> While the "plant" sprouts, emerges, and expands into the open, it simultaneously goes back into its roots, insofar as it plants them firmly into the closed ground and thus takes its stand. The act of self-unfolding emergence is inherently a going-back-into-itself. This kind of becoming present is φύσις. But it must not be thought of as a kind of built-in "motor" that drives something, nor as an "organiser" on hand somewhere, directing the thing. Nonetheless, we might be tempted to fall back on the notion that φύσει-determined beings could be a kind that *make themselves.* So easily and spontaneously does this idea suggest itself that it has become normative for the interpretation of living nature in particular [...]. No doubt a good deal of time has yet to pass before we learn to see that [this] idea ... is a purely modern, mechanistic-technological concept, according to which "growing things" are interpreted as artefacts that make themselves (Heidegger, 1998a, p. 195).

Heidegger's second point of departure from Wheeler is that he does not accept that a mechanical understanding of organism can be saved by positing additional supra-mechanical (non-reductive) capacities. Heidegger criticises this approach as failing to grasp the "proper essence of the living being":

> We must attempt to make biology and zoology recognise that organs are not merely instruments and that the organism is not merely a machine. This implies that the organism is something more, something over and above the machine. Yet the task is

> surely redundant because, either explicitly or implicitly, this is already recognised in the field of biology. But the fact *that* this is the case, and the *manner* in which it is the case, is precisely what is so fateful. Why? Because this recognition of a supra-mechanical moment actually appears to do justice to the proper essence of the living being. However, it does so in such a way that the initial approach is thereby sanctioned rather than overcome and is taken up into the fundamental determination of life. Here it returns in an even more virulent form to distort the original theory of the essence of life even further or to tempt it into introducing certain supra-mechanical forces (as in vitalism) (Heidegger, 1995, p. 217).

Here the fateful step that Heidegger takes issue with is the initial approach of taking up a mechanical understanding of the living being. He sees the idea of supra-mechanical capacities as sanctioning this initially misguided approach rather than overcoming it. The point we are arguing now is that Wheeler's non-reductive cognitive science has taken up just such a mechanistic conception of Dasein, founded on physical and functional subagential processes, and attempted to save it in just the same way, by proposing supra-mechanical phenomena that emerge from a physico-chemical mechanistic foundation. This means that Heidegger and Wheeler disagree about the fundamental task and starting point for the scientific conceptualisation of living being. For Heidegger the task is to develop a projection that does justice to the "proper essence of the living being" by *not* starting from a position that initially understands life as a physical/mechanical process. For Wheeler, the situation is reversed, he *starts* from a physical/mechanical understanding in order to make "it intelligible to us how it is that unmysterious causal processes (such as those realised in brains) can give rise to psychological phenomena that are genuinely constitutive of agency and cognition" (Wheeler, 2005, p. 127).

A.9 The Essential Difference

The *essential* difference between Heidegger and Wheeler is over how best to conceptualise the being of human agency. Wheeler remains in the third-person, objectifying conceptualisation of modern science, that takes everything, *initially,* to be present-at-hand. This conceptualisation, as Dennett observes, involves a basic de-essentialisation of whatever is objectified (Dennett, 1991, p. 421). As a result, for example, we have to develop the concept of qualia to express what has been left out of the third-person objectification of perceptual experience. Similarly, the essential

moments of what it means to be conscious, to have intentionality, to exist as being-in-the-world, are concealed and have to be explained or reintroduced or reduced *later.* This reintroduction of what was left out of the original projection becomes a major task of cognitive-scientific explanation. And yet, viewed from a Heideggerian perspective, this is a *self-created* task, one that arises out of a certain mode of projection, and not from out of the "matters themselves." For we are equally at liberty to acknowledge that there is an essential component to experience, intentionality, consciousness, that stands before us just as we stand before the tree in bloom. We can elect to project nature in such a way that this essentiality remains within the projection and so does not become a problem, as something that must be fitted in later, or explained away. In the field of biology, this is exactly the kind of projection that Heidegger was seeking.

For example, in his analysis of the animal as organism, Heidegger understands the eye as an *organ* that differs *essentially* from our conception of an instrument or machine. The basis of this difference is that the eye is *incorporated* into the organism, via the *capacity* to see, which itself is something essential that cannot be explained in terms of the mechanical functioning of the parts of the visual system:

> For example, we can understand *what the bee's eye achieves and its character as an organ* insofar as it is determined by *the bee's specific capacity for seeing.* The capacity for seeing on the other hand is not determined by the 'eye' of the bee and cannot be understood in this way. [. . .] The anatomical structure [of a bee's eye] can only provide some support for 'inferences' concerning the character of the bee's vision if and so long as we keep before us the essence of animality properly understood and the particular kind of animal being which belongs to a bee (Heidegger, 1995, p. 230).

This idea of the non-mechanistic essentiality of a capacity comes clearly to the fore when Heidegger asserts that the capacity *creates* the organ:

> [T]he capacity that the organ itself manifests does not belong to it as an organ, but rather . . . the organ belongs to the capacity. It is only in the sense that the capacity in each case creates its own organs that we can say that the organs are incorporated into the organism. For the capacity incorporates the organ into itself and retains the organ within itself. The organ remains an organ as long as it is retained within the organism (Heidegger, 1995, p. 227).

Here we should not think that the capacity *physically* creates the organ, rather that it creates the organ *as* an organ. This allows that artefacts such as cochlear implants can become incorporated into an organism, once they are taken into the service of a capacity to hear. In turn, Heidegger rejects the idea of the organism as a collection of capacities, and insists on the essential unity of the organism as a *capability*:

> The organism does not have capacities, that is, it is not an organism which is then additionally supplied with organs. Rather to say that 'the animal is organised' means that the animal is *rendered capable.* Being organised means *being capable.* And that implies that the animal's being is potentiality, namely the potentiality to articulate itself into capacities, that is, into those instinctual and subservient ways of remaining proper to itself. These capacities in turn possess the possibility of allowing certain organs to arise from them. *This capability articulating itself into capacities creating organs* characterises the *organism* as such (Heidegger, 1995, pp. 234–235).

A.10 Heideggerian Cognitive Science

It lies beyond the scope of the present work to further explain Heidegger's conceptualisation of animal being or to speculate on exactly how he would have projected human agency for contemporary cognitive science. The point of these examples is to show that he would *not* have started with a physicalist, scientific projection of being, within which a projection of Dasein as human agency is *secondarily* inserted at an agential level that floats in an unspecified (unreduced) way above the being of the physical. He would rather have started with Dasein, as he did with animal being, and understood the physical projection of science in terms of Dasein. In this way the existence and essentiality of Dasein would remain unproblematic because it is assumed from the outset.

A clear implication of this reasoning is that we cannot consider Wheeler's reconstruction of cognitive science as Heideggerian. It is rather the introduction of Heideggerian concepts into a framework that remains essentially naturalistic. On this basis we can understand Wheeler's book as a successful articulation of how such concepts can be incorporated into a embodied-embedded understanding of cognition. The difficulty arises when Wheeler argues that this incorporation reflects Heidegger's own view on the relationship between science and philosophy. We have argued that this view is mistaken, and that an essentially Heideggerian reconstruction of cognitive

science would involve a radical deconstruction of the physical/mechanistic conception of the brain from which contemporary thinking, whether orthodox or not, starts its journey.

Appendix B

The Phenomenological Negation of the Causal Closure of the Physical

[This paper was originally presented at the *Australasian Association for Philosophy Conference,* Wollongong University, July 1-6, 2012.]

B.1 Preliminaries

The aim of this paper is to provide a phenomenological demonstration that the causal closure of the physical is false. To begin I should like to clarify the notion of the physical I have in mind. As an example, consider the experience you are having now. I am assuming your eyes are open and you are seeing something. That something, taken in the ordinary sense, is the physical world, made up of physical spaces and objects, such as tables and chairs, and physical events involving such spaces and objects. In everyday conversation we know quite well what it means for something to be physical, and we can reliably indicate whether something is, in fact, physical. The concept is thereby used to distinguish between things and events that exist (in a way that can be publicly demonstrated), and the various experiences that can only be known to us privately or inwardly. It is this notion of the physical in contradistinction to entirely private, phenomenal, subjective experience that I shall be considering in this paper.

A paradigm example of an entirely subjective experience is my acquaintance with sensory colour, such as the redness of a particular shade of red. The demonstration that such experience is entirely subjective is my inability to express or indicate to anyone else what my experience of the particular red colour quality *qua* quality is (in itself). I can indicate the object whose objective

red colour is known to me on the basis of the colour quality. I can indicate the photons that are emitted from the object, the responses that occur in my retina, and the events occurring in my brain. But at no point in that chain of events can I indicate directly *to you* what it is that I am experiencing as the quality or essence of the colour of the object.

The best I can do is to ask you to inspect your entirely subjective experience of the colour of the object I am looking at, and work on the assumption that the colour quality you are experiencing is *similar* to the one I am experiencing (the likelihood that our qualitative experiences should concur exactly, given the differences in the responses of our respective visual systems, is not high). Perhaps, with a suitable brain-o-meter, we could observe that events of the same class are occurring in both our brains and, given that membership of this class determines the colour quality that is experienced, we could conclude we are both experiencing the same quality. The trouble here is that there is no conceivable experiment that could confirm, in general, that such and such a class of brain events determines such and such a colour experience (because my experience of colour quality is entirely subjective and so cannot be objectively corroborated).

Of course, I can simply decree that when a brain state of such and such a class obtains, then such and such a quality is experienced. The advantage (and disadvantage) of such a decree is that while no one can confirm it, no one can deny it either. I can even express my decree in the form of a psycho-physical law and claim a scientific provenance, because (after all), all scientific explanation must ground out in something one cannot explain, some ultimate brute fact or relation. Why not the brute fact of psycho-physical law? The problem here is that what we usually accept as scientific brute facts are facts whose effects we can observe objectively. So, for example, if we take Schrödinger's wave equation to be a brute fact, this is a brute fact whose validity can be tested by experimental observation. In contrast, the psycho-physical law that such and such a class of brain states determines such and such a qualitative experience cannot *in principle* be tested by experiment. What am I to say? "Yes I am experiencing quality x now"? But, as Wittgenstein pointed out, x cannot be defined. I have no precise idea what you mean by x and you have no precise idea what I mean by x. I cannot even verify that you are experiencing what-I-call-colour and not something that I would find incomprehensibly unfamiliar. I can only see *my* beetle, the one you can never see.[1] If we are to take Popper's maxim that a theory is only scientific if it

[1] Wittgenstein's point was that sensations (phenomenal qualities or beetles in boxes) cannot be meaningfully referred to as thing-objects within a standard language-game (Wittgenstein, 2001, p. 85, Section 293).

can be falsified, then the notion of psycho-physical law is *not* scientific.[2] Its function is rather to uphold a certain metaphysical position concerning the determinability of phenomenal experience.

So, for the purposes of this paper, the physical is the non-phenomenal, and the phenomenal is the pure quality of subjectivity, that which can only be known on the basis of direct or immediate experience. Given this category of the physical, we can now define the causal closure of the physical as the principle that 'every physical event is determined, in so far as it is determined at all, by preceding physical conditions and laws' (Montero & Papineau, 2005, p. 233). Causal closure therefore denies any *independent* causal efficacy to phenomenal experience. Accordingly, my experience of the quality of redness can enjoy a kind of epiphenomenal existence, but it can in no way, in and of itself, cause another physical event to occur. If it appears that such events do occur, such as my saying that "I am having a phenomenal red experience now," it must be that some underlying physical state has determined that utterance, just as some related physical state has determined my experience of the red quality.

In the remainder of the paper, I shall attempt to show that this notion of causal closure is false. The aim is to arrive at the insight that causal closure is false by means of a *direct seeing* that contains its own self-evidence. How this is to be achieved will become clearer as we continue.

B.2 The Paradox of Phenomenal Judgment

I think the most comprehensive and influential defence of the principle of causal closure is that provided by David Chalmers in his account of the paradox of phenomenal judgment.[3] The paradox arises from having already accepted the fundamental premise that reality is ontologically divided into the objectively physical and the subjectively phenomenal on the basis that the objectively physical is causally closed. Given this division, the ability to make correct judgments concerning subjectively phenomenal experiences appears paradoxical because phenomenal judgments are expressed as physical events, and yet the experiences of phenomenal quality about which we judge are not supposed to have any independent effects on physical events. Considered counterfactually, this amounts to supposing that speech acts concerning phenomenal experience will unfold in just the same way, *whether or not* there is any accompanying phenomenal experi-

[2] See (Popper, 1959/2002).

[3] The defence of causal closure is first laid out in Chapter 5 of *The Conscious Mind* (Chalmers, 1996) and then elaborated in subsequent papers that are now amalgamated in Chapters 8 and 9 of *The Character of Consciousness* (Chalmers, 2010a, 2010b).

ence.

Chalmers' answer to the paradox is to introduce the notion of *pure phenomenal concepts.* These concepts are physically instantiated in the brain, but bear no reference or relation to any objectively physical entity. Instead they refer to pure phenomenal qualities that are demonstrated directly in an immediate phenomenal experience. For example, I could be looking at a green leaf on a plant in front of me. Firstly, I grasp my experience of a patch of uniform greenness on the leaf as something distinct from my experience of the objectively physical leaf. I can indicate the experience as *this* green quality, meaning the colour quality I am experiencing now, subjectively, and only for as long as I am paying attention to the actual quality of that colour. An objectively physical account of colour perception would associate this subjective experience with certain processes occurring in my objectively physical visual system. However, the pure phenomenal concept does not refer to these physical processes either. It refers only to the experience itself, as a phenomenal experience.

Chalmers' argument concerning the paradox of phenomenal judgment contends that my consciousness of phenomenal quality makes no difference to the physical functioning of my brain or my speech behaviour. His idea is that the physical events in my brain that correspond to my forming a pure phenomenal concept, and uttering a judgment employing that concept, are determined (as far as they are determined), according to causal closure, i.e. by the preceding physical events and the physical laws that govern them. My experience of phenomenal quality is something additional that accompanies the physical formation of a pure phenomenal concept, and *constitutes* the content of that concept. The pure phenomenal content of the concept does not *cause* the concept to be formed, and thereby does not violate the principle of causal closure.

B.3 The Acquisition of Phenomenal Concepts

The problem with Chalmers' account is that it fails to resolve the paradox of phenomenal judgment. It rather shows how it is possible for a physically determined, unconscious entity to mimic a certain kind of human behaviour. What is not addressed is how such an unconscious entity could acquire the ability to wield pure phenomenal concepts in the first place. Instead, we are introduced to a fully formed brain, one that already possesses such abilities, and we are shown how this brain may continue to function in the absence of consciousness, in such a way as to utter judgments concerning pure phenomenal concepts that would satisfy a Turing test.

However, Chalmers offers an account of how an unconscious entity could acquire phenomenal concepts in *The Conscious Mind.* There he uses the notion of an information space and a processing system that finds itself within that information space, to explain how such a system could become puzzled about its experience:

> The crucial feature here is that when the system perceives a red object, central processes do not have direct access to the object itself, and they do not have direct access to the physical processes underlying perception. *All* that these processes have access to is the color information itself, which is merely a location in a three-dimensional information space.
>
> . . . Indeed, as far as central processing is concerned, it simply *finds itself* in a location in this space. The system is able to make distinctions, and it *knows* it is able to make distinctions, but it has no idea how it does it. We would expect after a while that it could come to *label* the various locations it is thrown into – "red," "green," and the like – and that it would be able to know just which state it is in at a given time. But when asked just *how* it knows, there is nothing it can say, over and above "I just know, directly." If one asks it, "What is the difference between these states?" it has no answer to give beyond "They're just different," or "This is one of *those,*" or "This one is *red* and that one is *green.*" When pressed as to what that means, the system has nothing left to say but "They're just different, qualitatively."
>
> . . . Given this kind of direct access to information states, then, it is natural to expect the system to use the language of "experience" and "quality" to describe its own cognitive point of view on perception. And it is unsurprising that all this will seem quite strange to the system: these immediately known, ineffable states, which seem so central to its access to the world but which are so hard to pin down. Indeed, it is natural to suppose that this would seem odd to the system in the same sort of way that consciousness seems odd to us.
>
> So this is the beginning of a potential reductive explanation of our judgments about consciousness: these judgments arise because our processing system is thrust into locations in information space, with direct access to those locations but to nothing else. The direct knowledge will strike the system as a brute "quality": it knows that the states are different, but cannot articulate this beyond saying, in effect, "one

> of *those*." This immediate access to brute differences leads to judgments about the mysterious primitive nature of these qualities, about the impossibility of explicating them in more basic terms, and to many of the other judgments that we often make about conscious experience. (Chalmers, 1996, p. 290-291).

It is to be noted here that Chalmers gives no account of the acquisition of *pure* phenomenal concepts. All that is shown is how unconscious systems can come to report and reason about the various physical configurations in which they find themselves. Chalmers' crucial assumption is that the central processes in his unconscious system 'do not have direct access to the physical processes underlying perception.' The 'mystery' for this system does not concern the existence of pure phenomenal qualities – if it is unconscious we must assume that it has no experience of such qualities. The mystery is rather a consequence of the system not 'understanding' the principles of its own operation. This kind of mystery is easily cleared up. The system only needs access to information concerning the design, construction and operation of its own components. Like a disciple of Dennett and the Churchlands, it will then 'understand' that its previous talk of mysterious qualities was an illusion. It only *seemed* that way because it did not have enough information concerning the physical realisation of its information space. Given this information, the entity will immediately 'understand' that its 'experience' of red can be entirely explained in terms of physical changes occurring within its physical components that cause it to utter statements to the effect that something is red when it is placed in front of paradigmatically red objects. In Chalmers' terminology, the system's phenomenal concepts will all be *relational* phenomenal concepts. They are relational because the corresponding phenomenal qualities are determined in relation to something else, i.e. states of affairs in the physical world that act as causes both in the forming and deployment of the relational phenomenal concepts. There will be no question of an additional 'quality' that is 'experienced,' there will just be the dispositions to utter certain phrases in certain situations concerning the detection of certain physical stimulii.

The question that Chalmers' paradox of phenomenal judgment poses is how an unconscious entity could acquire a *pure* phenomenal concept, i.e. a concept that refers to a pure phenomenal quality that bears no relation with any physical state, property or process. We, as conscious entities, already have sufficient information concerning the operation of the physical brain to see that our phenomenal experiences have corresponding physical manifestations. But this knowledge has not caused us to discard our notion that there are experiential qualities that are not captured by physical descriptions of the functioning of the brain. The task for Chalmers, in order to resolve

the paradox of phenomenal judgment, while maintaining causal closure, is to explain how an unconscious entity, entirely on the basis of its own 'experience,' and knowing all the details of its own physical operation, could come the conclusion, like us, that there is something more to being conscious than it already knows on the basis of physical science.

B.4 The Pre-Understanding of Pure Phenomenal Quality

Chalmers' account assumes that a pure phenomenal concept spontaneously comes into play once I direct my attention onto a pure phenomenal quality that is immediately present in my phenomenal experience and attempt to form a direct phenomenal belief concerning that phenomenal quality. However, in order to direct my attention in this way, I must already have an idea of what it is that I will find (i.e. an intention, or directedness-toward). That is, I must already understand that there is such a 'thing' as pure phenomenal quality that corresponds to my pure phenomenal concept.

It is this *pre*-understanding that Chalmers takes for granted. In order for his account to work, he needs to explain how it is that an unconscious entity could come to *notice* that there is such a 'thing' as a non-relational, pure phenomenal quality, in the first place. And this is something that an unconscious entity cannot do, because the very thing that the unconscious entity is unconscious of is non-relational, pure phenomenal quality. *If* an unconscious entity were to report having an experience of pure phenomenal quality, we would have to conclude that there was some error in its construction, an error that could be traced to some malfunctioning physical structure, and consequently rectified.[4]

Chalmers attempted to show how just such an error could be made in his account of information spaces quoted earlier. However, the information space scenario assumed that the unconscious entity remains in ignorance concerning the physical realisation of its information space. Once that artificial barrier is removed, i.e. once the system is given the same access to the world as we have, then Chalmers' argument collapses. It collapses because the rational response of such a system, once in possession of all the relevant information, is to assert, with Dennett, that eliminative physicalism is true.

The problem for eliminative physicalism, as Strawson has so trenchantly pointed out,[5] is that if we know anything at all, we *at least* know there is phenomenal experience of pure phenomenal

[4] See (Elitzur, 2009) for a refutation of causal closure that shows why an argument that rests on the premise of a machine that falsely concludes it is conscious must fail.

[5] See (Strawson, 2008).

quality. The question is, how can we know this, if phenomenal experience has no independent causal effect on our behaviour? Again, Dennett is right: if phenomenal experience has no independent causal effect on our behaviour, then we can't know that there is such a thing as phenomenal experience. Therefore our idea of phenomenal experience is an illusion.[6]

Chalmers' argument does not touch this conclusion. The problem is not to explain how I can refer to phenomenal quality using a pure phenomenal concept in such a way that the phenomenal quality is not causally implicated in the explanation. The problem is to explain how it is that I am able to maintain the *conviction* that there is such a thing as pure phenomenal experience in the face of the physical evidence concerning the operation of my brain. It is only on the basis of this conviction that I am able to distinguish between a pure and a relational phenomenal concept in the first place. For example, let us assume I am in possession of a relational phenomenal concept that is formed as I demonstrate a particular shade of green to myself, e.g. "*this* shade of green." My concept refers to the phenomenal experience occurring in me as a result of looking at a green leaf in front of me and so is *related* to my act of demonstration. It is quite conceivable that Chalmers' unconscious system could form a structure within itself that corresponds to this concept. When questioned further it would identify that phenomenal experience with certain events that the light emitted from the leaf causes to occur within its components, just as I can refer my phenomenal experience to certain events that the light causes to occur in my brain. The crucial difference comes when I ask the system to distinguish between its phenomenal experience of green and the corresponding activity of its components. It will have to say there that they are one and the same thing. But what about me? What is it that allows me think that my phenomenal colour experience is something more than the activity occurring in my brain? If my phenomenal experience is causally determined by the operation of physical law, then that experience cannot reach out of its causal dependency and start independently influencing my neurons. I may be passively conscious of the greenness of my phenomenal experience, but, according to causal closure, that passive consciousness can have no independent effect on the operation of my brain. So, despite my having the experience, I will be unable to form any thought that corresponds to or even registers the experience as anything more than the physical realisation of that experience. I will be necessarily mute on the subject, not just outwardly, I will be unable to even *think* that there is anything more to my experience than the physical functioning of my brain. I will be trapped in my relational phenomenal concepts, having my pure phenomenal experience, but powerless to

[6] See (Dennett, 1991, pp. 369-411).

form concepts with which to even indicate that such experience is occurring.

And yet, when I contemplate my phenomenal experience of the greenness of the leaf, I do form the conviction that there is something more to the experience than can be explained in objectively physical terms. If I examine this carefully, I do not form the conviction on the basis of some form of induction from experience. I know immediately, from within the experience itself. I cannot offer any *positive* justification, I immediately *see* that this is the case.

B.5 Phenomenal Concepts and their Reference

In the context of Chalmers' account, phenomenological 'seeing' is the act whereby I become aware of the pure phenomenal content of a pure phenomenal concept. This awareness occurs while I am attending to the colour of a suitably coloured object and ceases as soon as I stop so attending. The concept itself is unusual. Firstly, it only 'exists' for as long as I am engaging in a correct demonstration, and secondly, it refers to something peculiar: a pure phenomenal quality. It is this pure phenomenal quality that is (supposedly) 'seen.'

To examine this further, we shall require a model of what it means for something to be a concept, and what it means for that concept to refer to something. I shall take it that in ordinary experience, we encounter the world on the basis of a certain *perceptual intentionality.* That intentionality structures our present experience of the world both according to the actual data streaming through our sensory receptors and according to our past experience (insofar as we recognise the world as something familiar). For example, if I consider a perception of a coloured object in front of me now, I can say that my perceptual intentionality intends the object as a physical object, located in physical space, having surfaces coloured in such and such a way, and that the data streaming through my sensory receptors is confirming that intentionality. In addition, my recognising the object as, for instance, a pen, and my classifying the pen as being blue, involves certain pre-existing abilities that distinguish pens from pencils, blue things from red things, and so on. I take the possession of such abilities to indicate the presence of a *concept* of the thing or property that is successfully distinguished.

The pen itself, I take to be associated with a certain *intentional object.* This intentional object is what I am referring to when I think and speak of the pen. If, in fact, I am perceiving a physical pen that corresponds to this intentional object, then that physical pen is *identical* to the intentional object. My perceptual intending of the pen can be thought of as an expectation (instantiated by

physical processes in my brain) of a certain abstract form (instantiated within the stream of data arriving from the sensory receptors). If that form is detected, then I experience the identity of the form of my intending with a form of the world, i.e. I experience the conviction that the thing I am intending is really 'there' in the world.

However, I can also think and speak of things that do not exist in the way a physical pen exists (such as Sherlock Holmes and Doctor Watson) and I can *imaginatively* refer to such intentional objects, or I can refer to things as having been in the past (in memory), or I can expect the appearance of things in the future (in anticipation), or I can refer to entities I hypothesise may exist, and so on. In addition, I can intend or refer to *properties* of intentional objects, such as my pen being blue, and to *classes* of intentional object that share a common set of properties, such as blue pens, and to *states of affairs* involving relations between groups of intentional objects, such as the blue pen being on the table, and to *events* that involve changes in intentional objects, such as the blue pen falling off the table, and so on.

In speaking and thinking of these various kinds of intentional object, and of the properties, classes and relations that obtain between them, I am employing *concepts.* According to this model, it is the concept that refers, and its reference, in the first instance, is to an intentional object, or to the properties, classes and relations that obtain between such objects. As already mentioned, if a particular intentionality corresponds with the world (in a way that is inter-subjectively agreed within a community of language users) then any correctly formed concept that refers to that intentionality, *ipso facto* refers to the world itself. Finally, I take it that my intending of intentional objects (and the properties, classes and relations that obtain between them), and my use of concepts that refer to these intentional acts, are associated with and depend upon (i.e. supervene on) certain physical processes occurring in my brain.

The point here is not to argue whether such an account of reference is going to work in all cases, it is to provide a framework within which we can make clear the kinds of problems that are involved in trying to speak of phenomenal quality and how these problems are covered over if we do not remain vigilant in the way we use language and in the way we demonstrate or gain access to pure phenomenal experience.

Firstly, it is necessary to examine how colour concepts work in ordinary language. For example, how would you demonstrate to yourself the quality of the colour blue? Perhaps you would look around and point to the first blue object you could find and say (inwardly) "*this* is blue." Or perhaps there are no blue objects present, so instead you remember a blue object, or you imagine a

patch of blue. In each case you must look at, or think of, or imagine 'something,' i.e. you have to form an intention, and having formed the intention, you require the presence of a corresponding intentional object in order to have the fulfilment of a present experience of blueness. Even the imagination of a pure field of blue colour is the imagination of an intentional object: it is intended *as* something imagined, it has a certain imaginary extension, it either has borders, or it fades into indistinctness. Whereas blueness itself, as a pure phenomenal quality, has no spatial delineation. Similarly, if I were to experience a *sensation* of blue, then the sensation would be *another kind* of intentional object. For example, I can produce the afterimage of a colour by closing my eyes after looking at a bright light. Such an experience can be intended as a sensation, i.e. as the event that appears to be sensating in my eyes (like the tingling sensation that can be experienced in the hands). In the case of the afterimage, my intentional object is the afterimage sensation, and it is the *sensation* that has the property of being coloured, just as, when I look at the pen, it is the *pen* that has the property of being blue.

In these examples, the *quality* of blue is made present via the intending of an intentional object that has the (objective) *property* of being blue. The fact of an intentional object's being blue is a relational property that depends on the object and brings into play our ordinary, relational, public concept of colour. This concept is relational because the colour reference can be distinguished (in principle) in relation to the configuration of the physical world. So, for example, when I imagine a blue pen, certain areas of my brain become active that also become active when I perceive a blue physical object. And so on. In short, the property of something's being called or distinguished as blue is determined by the physics, and our correct use of the normal concept of blue can finally only be explained in terms of a set of suitably described physical entities and behaviours. The same goes for our ability to make fine distinctions between different shades of colour.

We shall call this concept of colour an *objective relational* concept because it refers to colour as a relational property of an intentional object that can be explicated physically. In contrast, when we consider a pure phenomenal concept, we are unable to give a physical account of how it is formed and to what it refers. To make this clear, consider again the case of my looking at a blue pen. If we follow the physical story, we can see how the presence of the pen activates an objective relational concept that enables me to conceive that there is a blue pen in front of me now. Following Chalmers, I can form the intention to highlight a particular uniform shade of colour on the surface of the pen, and form a concept that enables me to indicate the colour as *this* colour. Such an action can still be understood in terms of my forming an objective relational concept

that refers to a property of an intentional object. The feature that has gone unmentioned, is that the demonstration of the colour of a certain area on the surface of the pen is accompanied by an experience of that colour having a certain quality of blueness. It is this quality, that, so to speak, hovers above the physicality of the pen, the photons, and the physical events occurring in my brain, and that remains unaccounted for in the physical account of the reference of my objective relational concept.

B.6 Phenomenological 'Seeing'

The point of the preceding discussion is to understand that in using a pure phenomenal concept, and in thinking generally about such concepts and their referents, we are not dealing with concepts as they are usually understood and employed. Normal concepts (objective relational concepts) have a reference that can be fixed objectively in relation to the physical world. A pure phenomenal concept has no such reference. That means, if you employ a pure phenomenal concept as if it were a normal objective relational concept, you misuse it, and speak a form of nonsense. As Chalmers has already described, employing a pure phenomenal concept requires a *bringing to presence* of the actual quality. Unless you are able to *explicitly* demonstrate a *particular* pure phenomenal quality to yourself in this way, your *general* use of the concept will remain ungrounded. It is not enough to defer to another's use of the concept, e.g. by saying when I refer to a pure phenomenal quality, I am referring to what you are referring to. This works for objective relational concepts, because there is an objective fact of the matter that can (in principle) be verified. Whereas, when I refer to a pure phenomenal quality, I am referring precisely to that subjective quality known immediately to me. Whether you also experience the same quality is something I cannot verify. To repeat: if you are unable to bring such an experience of pure phenomenal quality to explicit consciousness, then you are also unable to correctly use any language that refers to pure phenomenal quality.

This is not a minor issue. It is clear that *many* philosophers who publish papers discussing phenomenal experience have been unable to bring an experience of pure phenomenal quality to explicit consciousness. Instead, the experience remains implicit, covered over by a sophisticated preunderstanding that determines the supposed referent of a pure phenomenal concept in advance. In order to achieve an *explicit* experience of phenomenal quality, one has to suspend, or bracket, any concept that purports to already understand what it is that is experienced. This is not a matter

of putting aside one concept and replacing it with another. As we have already demonstrated, a pure phenomenal quality cannot be represented by means of an objective relational concept, no matter how sophisticated. It is just not that kind of 'thing,' i.e. it is not a property of an intentional object.

This brings us to the phenomenological demonstration itself, i.e. the enactment of a phenomenological 'seeing' of pure phenomenal quality. Such explicit demonstration requires the creation of a discontinuity in the normal, more or less continuous process of intending the intentional objectivity that constitutes our normal waking experience of being in the world. This can be considered as a kind of limited or partial phenomenological epochē, enacted specifically in relation to a demonstration of pure phenomenal quality.[7] What is suspended is the everyday conceptualising intentionality that prefigures what you are about to experience on the basis of what you have previously experienced. Such prefiguring is habitual, natural and necessary. However, it is only on the basis of abstaining from such conceptualisation that the essence of colour as something non-conceptual can emerge. Otherwise the understanding remains bound to the intentional objectivities that the qualities exhibit, and exhibit in such a way that the constituting function of phenomenal quality within consciousness remains covered over. This is not a matter of positively attempting to conceive of phenomenal quality. The *fact* is that the pure phenomenal quality is already present, and not in a hidden way. The only obstacle to encountering this quality *directly* is having already understood the experience according an objective relational conceptuality (one can only repeat this in different ways).

Despite the preamble, what is being indicated is essentially simple. You find a coloured object in front of you. You look at the colour of the object. You stop interpreting your experience of this colour as being the colour of the object in front of you, or as being the colour of any 'thing'. You allow the colour to emerge as the colour that it is. You look at the colour experience itself, its quality, its essence. You look at what it is that makes your experience of blue *blue.* You remember that these words are indications only, that we cannot refer to this experience using ordinary objective relational concepts. And yet, it is only once the pure phenomenal quality is grasped as the quality that it is, that the language being used here will become intelligible.

[7]The phenomenological epochē intended here is related to Husserl's epochē of the objective sciences (Husserl, 1970/1992, pp. 135-137), except we are bracketing the objective relational conceptualisation of colour only, rather than bracketing 'all objective theoretical interests.' However, in practice, such a bracketing of colour experience also requires a bracketing of all objective scientific understandings that would, from the outset, deny the possibility of a direct intuition of phenomenal colour.

Finally, if and when you succeed in so suspending your normal objective relational interpretation of colour, you will discover something that you already knew (implicitly), something that you cannot directly express. It almost appears as if you have discovered nothing, or at least nothing worth mentioning. But, in *reality,* this is a pivotal event: you have discovered a phenomenon that lies outside the boundaries of physicality.

From here on we are speaking of the demonstration of pure phenomenal quality from *within* a phenomenological epochē, i.e. having already suspended the normal intentionality that only sees the coloured object that the quality of the colour presents. The task is to find a way to bring knowledge of this demonstration to explicit consciousness. Such "finding a way" is not prefigured, it is an *agreement* that must be reached before we attempt to speak of the experience. This agreement *is* the enacting of a phenomenological epochē. Given such an epochē, I can say that I now 'know' the quality 'directly', and not in the sense that I know an ordinary objectivity by means of an objective relational concept. There is no objectification that puts the quality 'over there' or as standing-against a subject who perceives or conceives the quality. It is rather that the phenomenal quality is known as the very means whereby an objective intentionality can become an object of conscious experience in the first place.

As Heidegger said (in a different context), we have 'let that which shows itself be seen from itself in the very way in which it shows itself from itself' (Heidegger, 1962/2008, p. 58). There are no 'hidden sides.' The quality as quality is completely revealed in a way that cannot be doubted. Doubt can only arise in relation to a concept that refers outside itself.[8] Whereas, my knowledge of pure phenomenal quality does not refer to anything outside the pure phenomenal quality itself - no 'thing' is the bearer of this quality and it does not exist as something separate from my knowledge of it.

This is not a question of correctly naming something. The pure quality has no name. To think that way is to confuse the quality with the objective relational colour instantiated in physical world. What fixes the reference of the colour in language are the physical conditions (my brain, the photons). We find no reason why we experience *this* quality rather than another, because all reasons are finally referred to the physical world and the physical world leaves the quality of the colour undetermined. From such a point of view *it doesn't matter* what quality I experience, because whether my blue is your green makes no difference to our use of language.

However, the fact is that I 'know' what blueness is. This knowing is a direct intuition. Unless

[8]Chalmers makes a similar point in his account of acquaintance (Chalmers, 2010b, pp. 285-294).

it is experienced as such, one has remained outside the epochē, one is *thinking* of the quality of blue using an objectifying concept that is a mere *sound.* Such thinking carries on regardless. Despite allowing there is such a thing as 'what it is like' to experience a colour quality, the experience itself is passed over, fixed in a certain framework of understanding that makes the quality itself inaccessible. Ordinary language achieves this covering over because it fixes experience in objective relational terms. There is no 'room' for phenomenal quality, because such language refers to intra-worldly entities, and phenomenal colour is *not* an intra-worldly entity. This is how phenomenal quality *escapes* ordinary language while at the same time remaining an essential moment within the experience of using language to refer to colour, in that our experience of phenomenal colour *expresses* what it is we mean when we refer to colour. It is *that-by-means-of-which colour is made known to us as colour.*

To think that this direct 'seeing' of phenomenal colour can become a premise within a philosophical argument is to again misunderstand the demonstration. Philosophical argument is public argument concerning the determination of objective relational concepts. One can never demonstrate the essence of colour ('prove' that it 'exists') via such a procedure. To think so is to have failed to make the demonstration explicit and to deliberately attempt to understand the demonstration in such terms is to make the direct seeing, the essence of phenomenal quality, disappear. One only need consider the case of Frank Jackson, and the literature on the 'knowledge argument' to 'see' where such an approach leads.[9]

What is 'seen' in the demonstration, itself reveals that we can only speak of pure phenomenal colour indirectly, i.e. via metaphor and via negation. We have to "circle round" the phenomenon and catch it "out of the corner of the eye," and then we have to say what it is *not,* i.e. it is not physical or intra-worldly or specifiable within objective relational concepts. This metaphorical circling and negation is a necessary consequence of the structure and limitations of ordinary objective relational language reference. However, to someone remaining within the stream of objective relational conceptualisation, it will appear that our inability to fix phenomenal quality in an adequate concept is a sign that we have not yet understood what it is, that we are dealing with some kind of fiction, or a phantom of the imagination. It is easy to become lost in such a line of thought. The answer is to stop and look again. The fact is there is direct intuition of phenomenal quality. It is true that phenomenal quality does not exist in the normal objective relational sense

[9] I am referring here to Jackson's intuition that colour experience involves a direct knowledge (Jackson, 1982) and his subsequent reconceptualisation of that intuition as an illusion (Jackson, 2003).

of the term. But phenomenal quality does 'exist' in the sense of being 'knowable.' Its existence *consists* in its being known. This existence is not something put together in the imagination. If phenomenal colour were something imagined then we should be free to vary the essence of colour in such a way as to imagine a new *kind* of colour, such as Terry Pratchett's octarine.[10] But this is not the case. The range of our phenomenal colour experience is strictly determined by the capacity of our nervous system to discriminate objective relational colour. We can, with Hume, imagine a *shade* of colour we have never actually seen, but we cannot imagine another *category* of colour.

Finally, we must draw a distinction between my *being* conscious of a pure phenomenal quality, and my *seeing* that I am conscious of a pure phenomenal quality. These are the two crucial moments within the unity of a conscious experience of phenomenal quality. The first moment, my being conscious of the quality, is the experience that Chalmers is indicating in his account of pure phenomenal concepts. This experience is a *pre-reflective* registering of the (non-physical) quality that is the *content* of a pure phenomenal concept. In Chalmers' account, as content, the pure phenomenal quality is experienced but has no independent causative agency. In this way the phenomenal character of consciousness is acknowledged without violating the principle of causal closure.

However, Chalmers misses the second moment of consciousness. In order to deploy a pure phenomenal concept I must already have direct intuition of pure phenomenal quality, i.e. I must be able to consciously demonstrate such a quality – because the quality only 'exists' as a quality insofar as I experience it. It is this demonstration that enables me to 'see' the distinction upon which my concept of pure phenomenal quality is founded.

Without such a 'seeing' demonstration I will have no conception that there is any such 'thing' as a pure phenomenal quality, even though I passively (pre-reflectively) experience pure phenomenal quality as the content of pre-existing natural language concepts. It is here that physical causation enters the account. For, my 'seeing' of pure phenomenal quality depends on something that is not physical, i.e. pure phenomenal quality itself. This quality, as a phenomenal quality, according to causal closure, cannot be the cause of any physical event. And yet, my consciousness of this quality, i.e. my direct experience of pure phenomenal quality, *is* the cause of a physical event. That physical event is my forming a new conceptual category that recognises the existence of pure phenomenal quality on the basis of my consciously experiencing that quality. In order

[10] See (Pratchett, 1985).

for this to happen, I not only have to pre-reflectively experience a pure phenomenal quality, I have to *see* that I am having such an experience. It is this *seeing* that is crucial for the negation of causal closure: it is where the direct intuition of phenomenal quality crosses over from being a passive (implicit) experiencing to being an explicit knowledge that makes a difference in the physical world. The essential point is that what is brought to explicit consciousness cannot itself be identified with the action of a physical process, i.e. a physical process cannot 'see' or represent an experience of phenomenal quality *as* an experience of phenomenal quality, it can only represent phenomenal quality as a physically instantiated objectivity. In contrast, my direct 'seeing' of phenomenal quality is a self-reflection of consciousness upon a pure experience of being conscious. As such it is nothing physical. And yet it makes a physical difference to my subsequent behaviour. I can now conceptually distinguish the being of pure phenomenal quality. Therefore the causal closure of the physical is false.

It should be noted that this seeing of pure phenomenal quality is not something that can be accounted for in terms of existing higher-order thought or higher-order perception theories. Such theories assume that higher order brain processes are *caused* by preceding physical events occurring in the brain. Whereas my conscious seeing of a pure phenomenal quality is not a physical event. A higher order thought that is physically determined cannot think about a pure act of conscious seeing, because, as far as the physics is concerned, no such act has occurred.

B.7 The Validity of the Phenomenological Demonstration

Chalmers' account of phenomenal judgment is that the pure phenomenal content of a pure phenomenal concept does not cause our judgment that our colour experience has *this* pure phenomenal content. Our judgments are determined by the physics of the brain, and it just turns out (perhaps on the basis of psycho-physical law) that conscious entities, such as ourselves, have the experience of pure phenomenal content whenever we attend (correctly) to our immediate phenomenal experience. The contradiction in Chalmers' argument is that *if* the pure phenomenal content of a pure phenomenal concept is not able to independently cause any event in the brain, then there is no mechanism whereby we can be caused to *see* that there is any such content. Whereas we do see that there is such content, and our seeing has physical effects, viz. our thinking and speaking of our seeing.

Nevertheless, it is conceivable, on the basis of mimicry, that an unconscious entity could form

a structure corresponding to my thought that "I am seeing the pure phenomenal quality of this colour now." Such an entity could then produce phrases of the same form in just those situations where I assert the same thing. In fact, we do not need to imagine unconscious entities to illustrate this point. It is quite conceivable in everyday life that there are people who will assert that they are seeing the pure phenomenal quality of a colour when in fact they are looking at a coloured object and repeating the phrase out of habit and without consciously appropriating the experience, i.e. without suspending their normal conceptualising of experience so that the quality of the colour is 'seen' as something distinct in itself.

In the case of someone who has *never* demonstrated the pure phenomenal quality of a colour to themselves, but speaks of such experience, their speech will be meaningless. It is not enough, as it is in the case of everyday reference, to defer to the knowledge of others. Despite my being immediately acquainted with colour throughout my conscious life, unless I have distinguished the pure phenomenal quality as something non-physical, then my reference will remain attached to the objectively physical entity that I take to be causing my experience. That entity may be the coloured thing in front of me, or the particular brain state that I believe is causing my experience. I can attempt, by close questioning, to try and elicit whether someone has really grasped what it means to see a pure phenomenal quality, but it is still possible that they are responding according to what other people have said on the matter. Until we have access to a brain-o-meter that can detect a physical difference, we must acknowledge that the only person I can be certain has grasped this matter is *myself.*

To the person who has never consciously appropriated an experience of pure phenomenal quality, it will also appear quite conceivable that there is no such quality, and that all such talk has developed on the basis of mimicry and false belief. It is no good arguing that we could not have originally developed a concept of pure phenomenal quality unless there were such qualities. I can imagine many things that do not have any reality. The *only* ground for the assertion that there is more to our phenomenal experience than can be explained in objectively physical terms, is the immediate seeing of that experience. The essentially private and subjective nature of phenomenological experience means this cannot be demonstrated through a procedure of adversarial argument. There is nothing objective that can be pointed to as a standard of truth. Consequently, all disputes in this area finally resolve down to the question of whether the immediate evidence of a phenomenological demonstration is decisive.

B.8 The Phenomenological Consequences

To summarise, if we accept the division of reality into the objectively physical and the subjectively phenomenal, and also that the realm of the objectively physical is causally closed, then we have arrived at a contradiction. The contradiction concerns my conviction that there is more to phenomenal experience than can be explained in objectively physical terms. If causal closure were true, and the activities of my brain and body could not in any way be independently influenced by the phenomenal quality of my phenomenal experiences, it follows that I should be unable to form any notion of my having pure phenomenal experiences that are not completely identified with objectively physical states of the world. In that scenario, we would all agree with Dennett, and eliminative physicalism would be an obvious truth. Whereas the empirical evidence is that eliminative physicalism is not an obvious truth, i.e. *many* people have the conviction that there is more to experience than can be explained in objectively physical terms. If we accept the evidence of experience and thereby accept this conviction as true, its truth can only be justified on the basis of a direct knowledge of phenomenal experience. If there is direct knowledge of phenomenal experience, and I am able to state that knowledge, as I am doing now, it follows that my phenomenal experience, properly (consciously) demonstrated, has had an effect on my physical behaviour. That effect is direct and simple: I *see* that there is something more to my experience than can be explained in objectively physical terms. It is that very seeing that is the non-objectively-physical cause of certain objectively physical behaviours, viz. my forming the proposition that "there is more to experience than can be explained in physical terms" and my uttering of that proposition. If phenomenal experience can determine physical behaviour in this way, then the principle of causal closure is false.

The consequence of this result is to put the ontological foundations of objective physicalism into question. Epistemically, the division of reality into the objectively physical and the subjectively phenomenal has been extraordinarily successful in terms of the progress of the physical sciences. However, that success does not imply that reality is ontologically distinguished along the lines that physical science tacitly assumes. Our failure to give a coherent account of human consciousness in terms of the objective physicalist program is a clear indication that we may have the ontology wrong. If we relinquish the principle of causal closure and allow that my being conscious (in certain circumstances) has an influence on my behaviour, then it is no longer possible to clearly divide reality along the axis of objective physicalism. That axis distinguishes phenomenal

experience on the basis of its causal dependence on the presence of objectively physical processes. It is this dependence that gives the objectively physical its ontological primacy. In denying causal closure, we also deny the coherence of attempting to explain phenomenal experience in entirely objectively physical terms. Phenomenologically speaking, objective physicalism is an unverifiable hypothesis concerning the possible structure of reality, a hypothesis that is negated on the basis of my phenomenological seeing that there is more to experience than can be explained in objectively physical terms.

Appendix C

Hierarchical Temporal Intentionality

[This is the first part of a paper originally presented at the *18th Conference of the Association for the Scientific Study of Consciousness,* University of Queensland, July 16-19, 2014.]

In recent years, a new and unified understanding of the functioning of the neocortex has emerged. This understanding sees the neocortex as a hierarchically structured Bayesian prediction machine that perceives and acts according to a delicate interaction between direct inputs from the body and environment, and feedback within the brain concerning what it expects those inputs to be. This *hierarchical predictive coding* model provides an elegant account of how attention, perception, cognition and action can be understood as different aspects of a single process that aims to minimise prediction errors.

However, despite the name, predictive coding models are not directly concerned with predicting the future, but rather with predicting what is to happen *now,* on the basis of having encoded the relevant structure of past input. As such, the predictive coding paradigm retains a "snapshot" understanding of perception that overlooks, and therefore leaves unexplained, the temporal horizons of experience. These temporal horizons were first clearly identified in Edmund Husserl's investigations of the unified tripartite retention-primal impression-protention structure of temporal consciousness. Several recent attempts have been made to explain how such a tripartite structure could be realised within current understandings of neocortical processing, but, as yet, none have been convincing.

In this paper I introduce an alternative model of neocortical processing that extends hierarchical predictive coding by proposing that the entire neocortex is engaged in sequence learning. As I shall explain, this hierarchical temporal memory model provides a coherent mapping be-

tween processes occurring in the brain and the structures of temporal consciousness identified by Husserl. The paper also provides a phenomenological examination and re-interpretation of the meaning of the hierarchical temporal memory model. This re-interpretation takes both consciousness and neocortical functioning to be fundamentally structured in terms of intentionality. Hence the title of the paper: hierarchical temporal intentionality.

C.1 Introduction

I think it fair to say that while there is broad agreement within the cognitive science community that phenomenology has some role to play in a scientific understanding of consciousness, the precise nature of that role has yet to be demonstrated in practice. One reason for this is that neuroscientific models of the large-scale functioning of the brain have lacked the detail needed to directly correlate neuronal function with phenomenological descriptions of experience. Where such neural correlates of consciousness have been indicated, such as the association of conscious experience with the phase locked firing of large, distributed cell assemblies, they have remained quite general and speculative (Freeman, 2000). A second reason is a basic lack of agreement concerning how to engage in phenomenological enquiry in the first place. From the naturalistic perspective of neuroscience, phenomenology is a matter of obtaining detailed and accurate descriptions of first-person introspection on a presumed realm of subjective experience. From a philosophical phenomenological perspective, such a naturalistic treatment of experience is seen as pre-judging and covering over the very domain it seeks to understand. Behind these divisions, however, is the more unifying idea that a proper phenomenological description of the essential structures of conscious experience should also provide practical guidance in the neuroscientific search for the neural correlates of such experience, and conversely, that a better understanding of the functioning of the brain should provide an empirical foundation and check on the validity of the structures that phenomenological enquiry is able to uncover.

The basic thesis of this paper is that current predictive coding models (see (Clark, 2013) of the functioning of the human neocortex have reached a point where such a unifying dialogue between neuroscience and phenomenology can begin in earnest. The aim is not to argue this in theory, but to demonstrate it in practice. Firstly, we must become clear concerning our mode of access to phenomenological experience. Here I shall be taking Husserl as the principal guide, for the practical reason that the structures Husserl uncovered are the most obviously relevant to the

models we shall be considering. Of course, it is by no means settled which theory of neocortical function is correct, or even if we are near to possessing such a theory. So there is a necessary element of speculation here, and also a degree of circularity, i.e. the models are being chosen *because* they fit so well with the phenomenology. Nevertheless, I am not assuming these models are correct. I am primarily pointing out an interesting correlation, one that weighs in their favour, all else being equal.

The particular model I shall be considering is Jeff Hawkins' memory prediction framework (Hawkins & Blakeslee, 2004; George & Hawkins, 2005; Hawkins et al., 2009; George & Hawkins, 2009). As will become clear, this model incorporates many of the features of the contemporary predictive coding schemes, but takes a radically different approach concerning the function of the lateral connections that exist within cortical regions. For Hawkins, the basic function of the neocortex is to encode (spatio-temporal) sequences rather than simple (spatial) patterns. Although this may appear to be a small difference of emphasis, the effect is to transform the neocortex from a machine that predicts its own next input into a phenomenologically plausible neural foundation for temporal consciousness. It is here that the connection with Husserl becomes apparent. For both Hawkins and Husserl, temporality is a foundational characteristic of experience. From Hawkins' neuroscientific perspective, the entire neocortical hierarchy is seen as a machine that processes spatio-temporal sequences of neural excitation in order to extract the underlying invariant structure from the ceaseless flow of sensory input, and to use this structure both to perceive and to act on that flow. From Husserl's perspective, temporal experience is also a matter of perceiving a world of stable objects on the basis of an immanent flowing experience that itself constitutes the temporality of a retained past, an immediately sensed present and a protended future.

It is this Husserlian temporal structure of experience that Hawkins' model is able to explain in detail. Previous work has already attempted to build such a bridge between phenomenology and neuroscience, but has only vaguely indicated how Husserl's structures of retention, primal impression and protention may be physically realised. In Hawkins' model, the recursive structure of temporal experience can be read off the hierarchically structured sequences that are activated by sensory input and the associated stream of feedback that flows down the hierarchy. It is in this downward stream that correlates with the phenomenological concept of intentionality. Hence I term Hawkins' model *hierarchical temporal intentionality.* Hawkins himself termed his model hierarchical temporal *memory* on the understanding that the hierarchy learns and remembers temporal sequences. However, once Hawkins' model is understood phenomenologically, it becomes

clear that intentionality is the more appropriate term.

C.2 Intentionality and the Phenomenological Reduction

For Husserl, all acts of consciousness are to be understood within an overarching framework of *intentionality.* This framework takes the fundamental and distinguishing characteristic of an act of consciousness to be its directedness-towards something other than itself. Such directedness-towards is experienced in the form of intentional meaning-bestowing acts. For example, when I perceive a table, I am enacting a perceptual intention that holds within itself a certain sense or meaning. What this meaning intends is the *intentional object* of the intending act, i.e. it is that towards which the act points. In the case of perceiving a table, the intentional object is the physical table itself. I do not intend, for instance, a mental process or internal image of the table. My intention reaches beyond my particular sphere of consciousness: I mean *that* objective table *there,* the one enduring object, the same enduring object that you intend when you perceive the table.

At the same time, my intending is *conscious.* That means that my intending is *immediately given* to me, not on the basis of reflection, but on the basis of a primary and basic *revealing.* The having of the immediate knowledge or experience of this revealing is what it means to be conscious. I cannot make this immediate revealing of the present moment an object of conscious reflection, as it is the very means by which what is reflected is revealed. I can only reflectively objectify a revealing that has already occurred, in which case it is no longer an actual revealing. Nevertheless, in being conscious, such revealing is immediately and *pre-reflectively* known in the very experiencing of acts that are intentionally directed away from themselves.

It is within this overarching understanding of pre-reflectively conscious intentional acts that Husserl formulates his idea of the *phenomenological reduction.* Put in basic terms, the reduction is a suspension of my acceptance of, or belief in the existence of a spatio-temporal world lying outside the sphere of my immediate intentional consciousness. Such a suspension requires that I already understand the intentional objects of my intending consciousness as being *transcendent* of the acts that intend them. For example, again considering my perception of a table, the act (my perceiving) is something *immanent* to my individual stream of consciousness, i.e. something particular and unique, never to be repeated in exactly the same way, comprising a series of phases or moments, themselves unique and unrepeatable. In contrast, what I intend in the act (the table

itself) does not change along with my stream of sensory-perceptive act-phases. My intention points towards what is *meant* in the act, and what is meant in the act is something *beyond* the act, something that remains *constant* throughout the changing phases of the act.

This intending of the (transcendent) table includes my intending the table in the mode of being a real physical object existing in a real physical world. Ordinarily (in the natural attitude) I live *within* this intending of reality, and unproblematically perceive the table as existing in front of me in the world of all existing things. Only if something goes wrong with the unfolding of my perceptual experience do I become aware of this *passive* reality-bestowing intentionality (such as would happen if my hand were to pass through the table without resistance).

The enacting a phenomenological reduction involves grasping that my *entire* experience of being in the world is founded on a universal reality-bestowing intentionality that precedes all my particular intentions that this or that object is real. In recognising this universal world-intending intentionality, it becomes possible to "stand aside" from it. That does not mean denying the validity of the intending-as-real, but rather allowing it to operate, while at the same time becoming reflectively aware of its operation. And in order to reflect on and identify this intending, one also has to suspend one's going along with it. It is this suspension of the going along with the intending of the world as transcendentally real that produces the reduction.

The reduction is therefore not an act of will whereby I engage in a "what-if" thought experiment, such as my imagining that I am a brain-in-a-vat and that the external world (as I perceive it) does not exist. Such acts of imagination function within an intentional framework that still accepts the world as real. A true phenomenological reduction enacts an understanding of consciousness (as intentional and therefore reality-bestowing) that produces a basic shift in perspective whereby my usual (passive and unrecognised) acceptance of the existence of the world becomes a matter of enquiry for me. In becoming a matter of enquiry, that acceptance, although still operating at one level (as that upon which I reflect), becomes transformed (in reflection) into something questionable. Such a putting into question can only occur on the basis that my acceptance is already suspended. And yet it is only on the basis of this suspension that the acceptance can be seen as an acceptance in the first place – otherwise I live entirely "within" acceptance-intentions that intend the world as being real (and not hallucinated, or imagined, or remembered, or dreamt). This circularity makes the reduction problematical: for how can I suspend an acceptance that I can only grasp on the basis of its already having been suspended? The short answer, and one that will have to suffice within the confines of the current discussion, is that an authentic reduction is founded

on an insight or direct seeing that is transcendent of the acceptance that it comprehends.

C.3 Phenomenology and the Neural Correlates of Consciousness

The phenomenological reduction reveals a stream of absolutely given (immanent) conscious experience. This stream comprises all the phases of my individual intentional acts. In turn, these acts intend intentional objects that transcend the stream in which they are intended. The difference the reduction makes is that I no longer go along with the meaning-intention that connects these intentional objects with a transcendent objective reality. That means I no longer unquestioningly consider the table in front of me to be an actuality, i.e. a real physical object. It has become a table *phenomenon,* intended in the *mode* of being actually real. More fundamentally, I no longer take my perception of a table to be an act occurring in a stream of consciousness that is connected to a physical body. The entire physical universe, including my empirical self (that I intend as an entity living in that universe), has also become a phenomenon, merely intended in the mode of being a real actuality.

What Husserl gains from the reduction is a new field of enquiry, whose absolute being has been disentangled from questions concerning the being of the transcendent physical universe. By disconnecting phenomenological experience in this way, Husserl, like Descartes, was seeking a realm of absolutely indubitable experience that could act as an ultimate foundation for all the knowledge claims of the objective sciences. The success or failure of this endeavour is not the issue here. Our concern is to connect Husserl's account of the structure of conscious experience, developed *within* the phenomenological reduction, *back* to the transcendent physical reality of the brain and its functioning.

Husserl himself had no interest in such a referring back – the very point of the reduction was to put physical reality out of play in order to demonstrate how the objectivity of that reality is constituted on the basis of conscious intentionality. In contrast, our aim is to use the phenomenological reduction as a source of evidence in the evaluation of current theories of the functioning of the neocortex. As the reduction is not widely accepted as such a source of evidence, we should examine why a specifically Husserlian reduction should be expected to yield useful results. The answer has to do with certain assumptions concerning the connection between phenomenological experience and the so-called *neural correlates of consciousness.*

Firstly, I am assuming that there are certain as yet unidentified processes occurring in the

brain that correspond *immediately* with the content of my phenomenological experience. The correspondence is immediate in the sense that there is no time or communication involved, i.e. there is no question of waiting for the neurons that comprise a correlated process to signal each other. The current state of their signalling *is* what correlates to my immediate conscious experience. In contrast, I am not immediately conscious of what is occurring on the underside of my foot. If I were to inadvertently step on a small stone, then it would take time for the nervous system to transmit the effects of that contact to my brain. Only once those effects have reached my brain does anything concerning my having stepped on a stone reach my conscious awareness (assuming I did not sense the stone in another way). Similarly, I am assuming there are processes occurring in my brain that are necessary for consciousness, but that only affect the content of my experience by means of signalling some other brain process that *is* immediately correlated with my conscious experience.

Secondly, I am assuming that the structural form of my flowing conscious experience is *lawfully related* to the structural form of those processes occurring in my brain that are immediately correlated with my conscious experience. The form of this relation is what concerns us in the remainder of the paper. For now I am taking it that (in principle) there exists a mapping whereby the fundamental ideal relations that can be identified within the processes of immanent conscious experience (on the basis of phenomenological reflection) can also be identified within the brain processes that are correlated with that experience (on the basis of neuroscientific investigation).

This does *not* imply that the brain represents objects in the world using neural "codes" that somehow embody the structure of the objects they represent. The picture is rather that each episode of conscious experience, say an ongoing experience of perceiving a dog, is instantiated in a widely-distributed temporally-enduring process occurring within the brain. It is within this process that we expect to find a pattern of neural excitation that corresponds to my concept of a dog. This could be an abstract relation instantiated by the activity of a (possibly changing) population of neurons that consistently arises whenever I form an intention whose intentional object is a dog. We are *not* expecting this dog-concept-pattern, in itself, to resemble a dog (although we are not ruling it out). What we *are* expecting is that any such currently active dog-concept-pattern will function within a wider (enduring) brain process in such a way that we can map the structure of that functioning onto immanent experiential structures that are accessible to phenomenological reflection.

If this overall picture is correct, then phenomenology has a central role to play in the develop-

ment of a correct theory of brain function, not just in providing empirical observational data to test a given theory, but in guiding the process of theory formation in the first place. For, if certain processes in the brain are absolutely (i.e. immediately) given to consciousness, and there is a lawful mapping between the abstract structure inherent in those processes and the structure of conscious experience, it should be possible to characterise the structure of those physical processes, purely on the basis of a proper phenomenological analysis.

The crucial issue, of course, is to determine what constitutes a proper phenomenological analysis, i.e. to determine how the *relevant* experiential structures can be distinguished from those that are irrelevant. One cannot simply describe what one experiences. What one experiences is the world, the table, the dog, and so on. It is not the structure of the dog as a dog that is of interest. It is the structure of the experience of perceiving a dog that is of interest. But how do we separate the structure of the experience from the structure of what is experienced? Answer: by means of a phenomenological reduction. The reduction is central because it allows us to divide experience in such a way that it corresponds with our understanding of what is occurring in the brain. This can be summarised in the following two points:

Firstly, the reduction identifies as immanent just that aspect of experience that corresponds to the activity of the neural correlates of consciousness. Here the formal-relational content of what is absolutely immanently given, i.e. the unique moment to moment phases of my intentional experience, is directly correlated with the unique functioning of the underlying neural populations.

Secondly, the reduction, in suspending transcendent reference, is concerned with identifying just those deeper, temporally enduring structures (i.e. my meaning intentions) that are synthesised out of the data of my immanent moment to moment experience. These deeper structures correspond with the temporally enduring forms of organisation that are emerging in the same underlying physical processes that specify my immanent experience. It is the knowledge of these structures that we expect to guide our understanding of the high level functioning of the associated areas of the brain.

To be clear, the foregoing considerations are not intended as formal arguments. We are seeking a way forward in understanding how experience is related to brain function and we are taking Husserlian phenomenology as a starting point (from the side of experience). Whether it is a correct starting point is something to be evaluated on the basis of the neuroscientific evidence. To begin, we shall look in more detail at the kinds of structures that Husserl's investigations uncovered. We have already accepted that experience is structured in terms of an overarching intentionality,

but we have only loosely characterised this intentionality in terms of simple acts of perception. We have not considered the underlying temporal structure of these acts, and in so doing we have naïvely accepted a worldly conception of time. It is this underlying temporality that is the characteristic signature of all intentional acts. If we are correct in expecting a lawful mapping to exist between the basic structures discoverable in phenomenologically-reduced experience and the corresponding abstract structure of processes occurring in the brain, then we should also expect this underlying temporal structure to be a ubiquitous and salient feature of all brain processes that are immediately correlated with conscious intentional experience. For this reason (and others that will become clearer as we continue) we now turn our (phenomenologically-reduced) attention onto the question of *temporality.*

C.4 The Phenomenology of Internal Time Consciousness

One's entire intentional life is unified within the flowing structure of internal time consciousness. This structure remains invisible in the experience of external time, because that experience *intends* time as external, i.e. as the temporal dimension within which external events occur and obtain their objective time determinations. Within a phenomenological reduction, one can examine this time constituting intentionality as it operates. This first requires thematising the flowing nature of immanent experience, and recognising that the '*universe of subjective processes,* which are the "really inherent" consciousness-constituents of the transcendental ego, is a universe of compossibilities only in the universal *unity-form of the flux,* in which all particulars have their respective places as processes that flow within it' (Husserl, 1999, p. 75). Husserl is here employing *eidetic intuition* to uncover the universal form of experience. Such intuition proceeds on the basis of a free variation of possible experience such that the invariant or generalised abstract form of the variations is given as an '*intuitive and apodictic consciousness of something universal*' (Husserl, 1999, p. 71). Such apodictic consciousness is possible within a reduction, because the immanent experience in which the universal is discovered is itself absolutely given, and can therefore be subjected to an absolutely free variation.

It is within this moment to moment flowing of experience that I intend intentional objects that do not flow, that endure and remain self-identical, even though my very acts of intention are themselves a part of this stream – they too have their particular and never to be repeated phases, their beginnings and endings, their flowing off. My perceiving of the table is *particular,* it is a

unique event. My turning away and turning back towards the table initiates another, numerically distinct act of perception. And as for the acts, so for each phase of each perceptual act. Each is particular and unique, whereas the table that I intend is the same table, in both acts, and in each phase of the two acts, and the table that you intend, if you perceive the table, is the self-same table. So we ask how is it that we experience permanence within perpetual change?

Husserl addresses this question most thoroughly in his famous analysis of the immanent experience of hearing a tone:

> We now exclude all transcendent apprehension and positing and take the tone purely as a hyletic datum. It begins and ends; and after it has ended, its whole duration-unity, the unity of the whole process in which it begins and ends, "recedes" into the ever more distant past. In this sinking back, I still "hold onto it," have it in a "retention." And as long as the retention lasts, the tone has its own temporality; it is the same, its duration is the same. I can direct my attention to the way in which it is given. I am conscious of the tone and of the duration it fills in a continuity of "modes," in a "continual flow." At one point, one phase of this flow is called "consciousness of the commencing tone"; and in this phase I am conscious of the first time-point of the tone's duration in the mode of the now. The tone is given, that is, I am conscious of it as now. But I am conscious of it as now "as long as" any one of its phases is intended as now. However, if any temporal phase . . . is an actually present now (with the exception of the initial phase), then I am conscious of a continuity of phases as "immediately past" and of the whole extent of the temporal duration from the beginning-point up to the now-point as elapsed. I am not yet conscious of the remaining extent of the duration, however. When the final point is reached, I am conscious of this point itself as the now-point and of the whole duration as elapsed. . . . "Throughout" this whole flow of consciousness, one and the same flow is intended as enduring, as now enduring. "Beforehand" (in the event it was not expected), it is not intended. "Afterwards" it is "still" intended "for a time" in "retention" as having been; it can be held fast and stand or remain fixed in our regard. The whole extent of the tone's duration or "the" tone in its extension then stands before me as . . . a formation no longer vitally animated by the generative point of the now but continuously modified and sinking back into "emptiness." The modification of the whole extent, then, is . . . essentially identical with the modification that the elapsed part

> of the duration undergoes in the transition of consciousness to ever new productions during the time that the tone is actually present (Husserl, 1991, pp. 25-26).

Here, in examining the flowing off of a tone, I can discern an actually present experience of the past-as-retained. This retention is not a memory, as can be demonstrated by remembering the elapsing of the same tone. Such remembering, if it is to fully reproduce the original experience, will have to reproduce the retentional structure of the original experience, and so cannot explain that structure. Neither is the retained content experienced as being simultaneous with the immediately present (sensed) now-point of the tone, i.e. I do not experience a retained and a currently sensed tone content as a kind of chord. The past tone content is intended as past and retains a fixed temporal position within a succession of tone phases that remain 'spread out' (so to speak) within an all encompassing consciousness.

In such reflection on temporal consciousness, I can discern that each immanent 'now' of consciousness 'contains' content that is intended as being now (what Husserl terms the *primal impression*) and content that is intended as being past (what Husserl considers to be retentions of former primal impressions). This means I have *two* nows, the one is the all-encompassing now of my immediate and pre-reflective consciousness, and the other is the now-point that I intend within that consciousness, as distinct from the former nows that I retain and intend as being past.

In addition, Husserl identifies intentions within the immediate now of consciousness that intend the future. These future intentions, or *protentions,* embody my expectation of future primal impressions. As expectations of what has yet to occur, protentions do not have the concreteness of a retention, because they have yet to be *fulfilled* in the sensory experience of a primal impression. Husserl terms them *empty* intentions, not because they are without form, but because they lack such sensory fulfilment. Empty intentions are pure meanings, and, as such, are harder to identify in experience than already fulfilled retentions. This is not because protention is somehow less important than retention, or because it operates occasionally, or because it does not play an important role in the synthesis of experience. It is rather that protention is hard to identify *because* it is such a fundamental aspect of experience. As our later examination of the neuroscience will clarify, perceptual experience is to a large extent determined by what I *expect* to experience, i.e. by what I protend. If the flow of my protending corresponds unproblematically with what is being sensed then this protending remains invisible. It is only when my protending of the stream of primal impressions breaks down in some way that I become aware of my protending, i.e. I experience surprise.

Protention creates horizons of expectation within which my entire experiential life is situated. These horizons provide the context and the measure of experience. For example, it is only on the basis of the harmonious (temporal) unfolding of the horizontal structure of protention that I come to accept an intentional object of perception as being transcendently actual or real in the first place. It is this harmonious fulfilment of protention that determines the very acceptance that I suspend in a phenomenological reduction, and that reveals Macbeth's dagger to be an hallucination (I expect real daggers to be resistant to touch and not to float in the air). Every moment of perception is subject to this scrutiny of protention: each object must transform itself in my visual field according to certain specifiable rules of perspective, each step I take must produce the right sound, the right feeling of resistance, in the right temporal ordering, and so on.

If my protending breaks down in some way, such as my misperceiving a stick as a snake, then my protending changes and this change and my previous protending are also *retained* along with the associated primal impressions. For example, at one moment I am perceiving a snake (the snake is my intentional object) and I am protending how the snake will appear as I move closer, and this perceptual act, including what *it* retains, is itself retained. In the next moment, the pattern of the snake that I protended is *not* fulfilled, my primal (perceptual) impression becomes a perception of a stick, my protention becomes the protention of the stick, and this entire retention-primal impression-protention structure is itself retained. However, this retention also retains the previous moment, and that moment contains the previous protention (of a snake) and the previous retention (of a snake perception). In other words, the structure of retention is *recursive* as illustrated in the following adaptation of Husserl's famous diagram:

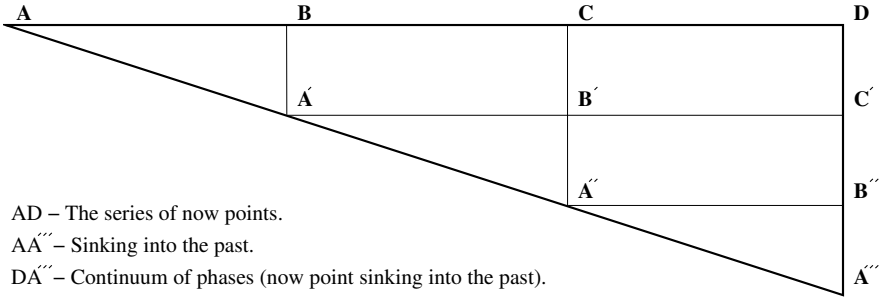

Figure C.1: Diagram representing four successive phases of consciousness, showing the initial primal impression A (on the left) sinking into the past and retained as A' as the new primal impression B arrives, then as B is replaced by C, A itself sinks further back and is retained as A'' as B is retained as B' and so on.

References

Brentano, F. (1995). *Psychology from an empirical standpoint* (A. C. Rancurello, D. B. Terrell, & L. L. McAlister, Trans.). New York: Routledge.

Chalmers, D. J. (1996). *The conscious mind.* Oxford: Oxford University Press.

Chalmers, D. J. (2010a). The content of phenomenal concepts. In D. J. Chalmers (Ed.), *The character of consciousness* (p. 251-275). Oxford: Oxford University Press.

Chalmers, D. J. (2010b). The epistemology of phenomenal belief. In D. J. Chalmers (Ed.), *The character of consciousness* (p. 277-304). Oxford: Oxford University Press.

Churchland, P. M. (1988). *Matter and consciousness: A contemporary introduction to the philosophy of mind*. Cambridge, Mass.: MIT Press.

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioural and Brain Sciences*, *36*, 181–253.

Dennett, D. C. (1987). *The intentional stance*. Cambridge, Mass.: The MIT Press.

Dennett, D. C. (1991). *Consciousness explained.* Boston, Mass.: Little, Brown and Company.

Descartes, R. (1997). *Key philosophical writings* (E. L. Haldane & G. R. T. Ross, Trans.). Ware, Hertfordshire: Wordsworth Editions Limited.

Elitzur, A. C. (2009). Consciousness makes a difference: A reluctant dualist's confession. In A. Batthyány & A. C. Elitzur (Eds.), *Irreducibly conscious: Selected papers on consciousness* (p. 43-72). Heidelberg: Universitätsverlag Winter.

Elton, M. (2003). *Daniel Dennett: Reconciling science and our self-conception*. Cambridge: Polity Press.

Fichte, J. G. (1889). *The science of knowledge* (A. E. Kroeger, Trans.). Ludgate Hill, London: Trübner and Co.

Fink, E. (1995). *Sixth Cartesian meditation: the idea of a transcendental theory of method* (R. Bruzina, Trans.). Bloomington: Indiana University Press.

Fodor, J. A. (1975). *The language of thought*. New York: Crowell.

Freeman, W. J. (2000). *How brains make up their minds*. New York: Columbia University Press.

Friston, K. (2010). The free energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*, 127–138.

George, D., & Hawkins, J. (2005). A hierarchical Bayesian model of invariant pattern recognition in the visual cortex. In *Proceedings of the international joint conference on neural networks*

(ijcnn-05) (pp. 1812–1817).

George, D., & Hawkins, J. (2009). Towards a mathematical theory of cortical micro-circuits. *PLoS Computational Biology*, *5*(10), e1000532.

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, *27*(3), 377–396.

Grush, R. (2005). Internal models and the construction of time: Generalizing from state estimation to trajectory estimation to address temporal features of perception, including temporal illusions. *Journal of Neural Engineering*, *2*(3), S209–S218.

Grush, R. (2006). How to, and how *not* to, bridge computational cognitive neuroscience and Husserlian phenomenology of time consciousness. *Synthese*, *153*, 417–450.

Hawkins, J., Ahmad, S., & Dubinsky, D. (2011). *Hierarchical temporal memory including HTM cortical learning algorithms* (Tech. Rep.). Palto Alto: Numenta Inc.

Hawkins, J., & Blakeslee, S. (2004). *On intelligence*. New York: Henry Holt.

Hawkins, J., George, D., & Niemasik, J. (2009). Sequence memory for prediction, inference and behaviour. *Philosophical Transactions of the Royal Society B*, *364*, 1203–1209.

Heidegger, M. (1962/2008). *Being and time* (J. Macquarrie & E. Robinson, Trans.). New York: Harper Perennial.

Heidegger, M. (1968/2004). *What is called thinking?* (J. G. Gray & F. Wieck, Trans.). New York: HarperPerrenial.

Heidegger, M. (1977a). The age of the world picture. In W. Lovitt (Ed. & Trans.), *The question concerning technology and other essays.* New York: Harper and Row.

Heidegger, M. (1977b). Science and reflection. In W. Lovitt (Ed. & Trans.), *The question concerning technology and other essays.* New York: Harper and Row.

Heidegger, M. (1992). *A history of the concept of time: Prolegomena* (T. Kisiel, Trans.). Bloomington: Indiana University Press.

Heidegger, M. (1995). *The fundamental concepts of metaphysics: World, finitude, solitude* (W. McNeill & N. Walker, Trans.). Bloomington: Indiana University Press.

Heidegger, M. (1998a). On the essence and concept of physis (T. Sheehan, Trans.). In W. McNeill (Ed.), *Pathmarks.* Cambridge: Cambridge University Press.

Heidegger, M. (1998b). On the essence of ground. In W. McNeill (Ed. & Trans.), *Pathmarks.* Cambridge: Cambridge University Press.

Heidegger, M. (2002). *On the essence of truth: on Plato's parable of the cave allegory and*

Theaetetus (T. Sadler, Trans.). London: The Continuum Publishing Company.

Hume, D. (1910). *An enquiry concerning human understanding*. New York: P. F. Collier and Son.

Husserl, E. (1970/1992). *The crisis of European sciences and transcendental phenomenology* (D. Carr, Trans.). Evanston, Illinois: Northwestern University Press.

Husserl, E. (1991). *On the phenomenology of the consciousness of internal time (1893–1917)* (J. B. Brough, Trans.). Dordrecht: Kluwer Academic Publishers.

Husserl, E. (1997). *Thing and space: Lectures of 1907* (R. Rojewicz, Trans.). Dordrecht: Kluwer Academic Publishers.

Husserl, E. (1999). *Cartesian meditations* (D. Cairns, Trans.). Dordrecht: Kluwer Academic Publishers.

Jackson, F. (1982). Epiphenomenal qualia. *The Philosophical Quarterly*, *32*(127), 127-136.

Jackson, F. (2003). Mind and illusion. In A. O'Hear (Ed.), *Minds and persons* (p. 251-272). Cambridge: Cambridge University Press.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in visual cortex. *Journal of the Optical Society of America A*, *20*(7), 1434–1448.

Lloyd, D. (2002). Functional MRI and the study of human consciousness. *Journal of Cognitive Neuroscience*, *14*(6), 818–831.

Lloyd, D. (2004). *Radiant cool: A novel theory of consciousness*. Cambridge, Massachusetts: The MIT Press.

McDowell, J. (1994a). The content of perceptual experience. *The Philosophical Quarterly*, *44*(175), 190–205.

McDowell, J. (1994b). *Mind and world*. Cambridge, Massachusetts: Harvard University Press.

Minsky, M. (1988). *The society of mind*. New York: Simon and Schuster, Inc.

Montero, B., & Papineau, D. (2005). A defence of the *via negativa* argument for physicalism. *Analysis*, *65*(3), 233–237.

Petitot, J., Varela, F. J., Pachoud, B., & Roy, J.-M. (1999). *Naturalizing phenomenology: Issues in contemporary phenomenology and cognitive science*. Stanford. CA: Stanford University Press.

Popper, K. (1959/2002). *The logic of scientific discovery*. New York: Routledge Classics.

Pratchett, T. (1985). *The colour of magic*. London: Corgi Books.

Rao, R. (2004). Bayesian computation in recurrent neural circuits. *Neural Computation*, *16*(1),

1–38.

Rao, R., & Ballard, D. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*, 79–87.

Sacks, O. (1985). *The man who mistook his wife for a hat*. London: Gerald Duckworth & Co. Ltd.

Schopenhauer, A. (1969 i). *The world as will and representation, Volume 1* (E. F. J. Payne, Trans.). New York: Dover Publications, Inc.

Schopenhauer, A. (1969 ii). *The world as will and representation, Volume 2* (E. F. J. Payne, Trans.). New York: Dover Publications, Inc.

Strawson, G. (2008). Real materialism. In G. Strawson (Ed.), *Real materialism and other essays* (p. 19-52). Oxford: Clarendon Press.

Thornton, J., Faichney, J., Blumenstein, M., & Hine, T. (2008). Character recognition using hierarchical vector quantization and temporal pooling. *AI 2006: Advances in Artificial Intelligence, Lecture Notes in Computer Science*, *5360*, 562-572.

Thornton, J., Gustafsson, T., Blumenstein, M., & Hine, T. (2006). Robust character recognition using a hierarchical Bayesian network. *AI 2006: Advances in Artificial Intelligence, Lecture Notes in Computer Science*, *4304*, 1259-1264.

Thornton, J., Main, L., & Srbic, A. (2012). Fixed frame temporal pooling. *AI 2012: Advances in Artificial Intelligence, Lecture Notes in Computer Science*, *7691*, 707–718.

Thornton, J., & Srbic, A. (2013). Spatial pooling for greyscale images. *International Journal of Machine Learning and Cybernetics*, *4*(3), 207–216.

van Gelder, T. (1996). Wooden iron? Husserlian phenomenology meets cognitive science. *Electronic Journal of Analytic Philosophy*, *4*.

Varela, F. J. (1999). The specious present: A neurophenomenology of time consciousness. In P. Jean, F. J. Varela, B. Pachoud, & J.-M. Roy (Eds.), *Naturalizing phenomenology: Issues in contemporary phenomenology and cognitive science* (pp. 266–329). Stanford, CA: Stanford University Press.

Varela, F. J., Thompson, E., & Rosch, E. (1993). *The embodied mind: cognitive science and human experience*. Cambridge, Mass.: MIT Press.

Wheeler, M. (2005). *Reconstructing the cognitive world: the next step*. Cambridge, Mass.: MIT Press.

Wittgenstein, L. (2001). *Philosophical investigations*. Oxford: Blackwell.