CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Being me: knowing-by-being, primary facts, and bodily selfhood
Authors	Parker, Robert J.
Publication date	2020
Original Citation	Parker, R. J. 2020. Being me: knowing-by-being, primary facts, and bodily selfhood. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2020, Robert J. Parker. - https://creativecommons.org/licenses/by-nc-nd/4.0/
Download date	2026-11-01 06:33:23
Item downloaded from	https://hdl.handle.net/10468/10871



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Ollscoil na hÉireann, Corcaigh

National University of Ireland, Cork



BEING ME

Knowing-by-being, Primary Facts, and Bodily Selfhood

Thesis presented by

ROBERT J. PARKER

for the degree of

Doctor of Philosophy

University College Cork

DEPARTMENT OF PHILOSOPHY

Head of Department: Professor Don Ross

Supervisor: Dr. Joel Walmsley

2020

CONTENTS

Declaration……5

Abstract……6

Introduction……8
(i) Self-referentiality……8
(ii) Preview of Chapters……16

Chapter 1. KNOWING-BY-BEING & PRIMARY FACTS……20
1.1 In the beginning was the deed……20
1.2 Knowing-by-being……22
1.3 Primary Facts……25
1.3.1 Consciousness……27
1.3.2 Independence……28
1.3.3 Intersubjectivity……30
1.3.4 Interdependence……31
1.3.5 Unity……33
1.3.6 Compositionality……34
1.3.7 Temporality……36
1.3.8 Judgment……37
1.3.9 Agency……39
1.3.10 Teleology……40
1.4 Elusiveness v Direct Awareness of Self……41

Chapter 2. PHENOMENOLOGY & THE METAPHYSICS OF SELF……47
2.1 Pre-reflective self-consciousness……48
2.2 The Phenomenology of self……50
2.3 From Phenomenology to Metaphysics……52
2.4 My Mineness……59
2.5 A Bodily Self……63
2.6 Knowing-by-being a bodily self……69

Chapter 3. PHYSICS, PHYSICALISM & CONSCIOUSNESS……72
3.1 Worldviews and a Theory of Everything……72
3.2 The scientific picture of the world……74
3.3 Physicalism……76
3.4 Physics and Mathematics……79
3.5 The Interpretation of Quantum Physics……84
3.6 The barrier of consciousness and the burden of proof……89
3.7 Escape routes……93
3.7.1 Watchful waiting……93
3.7.2 Embrace emergence……96
3.7.3 Favor fundamentality……101
3.8 Towards modifying metaphysics……104

Chapter 4. BIOLOGICAL SELVES & KNOWING-BY-BEING-ALIVE……106
4.1 Biological Individuals……106
4.2 Cells……107
4.3 Organisms……110
4.4 The Immune System……111
4.5 Autonomous Wholes……114
4.6 Hylomorphism……116
4.7 Knowing life by being alive……119

Chapter 5. SCIENTISM & THE METAPHYSICS OF BEING HERE……122
5.1 Modifying metaphysics……122
5.2 Scientism as metaphysics……123
5.3 Re-visioning the nature and role of science……129
5.4 Phenomenology and science……134
5.5 The 'problem' of consciousness……137
5.6 Dissolving the hard problem……139
5.7 The metaphysics of being here……142

Chapter 6. COUNTING SELVES……147
6.1 How many selves are there?……147
6.2 None……148
6.2.1 Buddhist No-Self……148
6.2.2 Relational Persons……153
6.2.3 Illusionism……159
6.3 One Self……167
6.4 Group Selves……170
6.5 Elephants, Octopuses, Cells & Selves……175

Chapter 7. WHY SELVES COUNT……183
7.1 Anthropic Analysis……183
7.2 The Weak Anthropic Principle……188
7.3 Explanatory Knowledge and Judgment……192

Chapter 8. THE PRIMACY OF BEING ME……198
8.1 Individual Conscious Subjects……198
8.1.1 The Universe and Me……198
8.1.2 Being a Conscious Subject……199
8.1.3 Being Me……201
8.2 Objectivity, Existential Primacy and Ontology……202
8.2.1 Objectivity……202
8.2.2 Existential Primacy……205
8.2.3 The Ontological Status of Consciousness……206
8.3 The view from here……208

Conclusion……210

References……213

DECLARATION

This is to certify that the work I am submitting is my own and has not been submitted for another degree, either at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism.

Robert J Parker.

ABSTRACT

In this dissertation I re-assert the significance of the ancient Greek aphorism - 'know thyself': I identify the individual human self as the starting point and ubiquitous preamble of all epistemology and ontology.

I argue that the primary feature of the way we find ourselves living is as an individual conscious bodily subject (which I express as 'being-me') and this is the locus or starting point (which I express as 'knowing-by-being') of all our possible knowledge.

Although we find ourselves as individual subjects of experience and action we are not isolated; rather we are always already with others as part of a wider world. I therefore propose a proto-ontology of 'me, others, and world' with corresponding interlocking modes of 'individuality, intersubjectivity, and interdependence'. I argue that our direct knowing-by-being of our selves reveals some further primary facts about ourselves in terms of unity, compositionality, temporality, judgment, agency, and teleology.

I acknowledge the phenomenological concept of pre-reflective self-awareness as an expression of the universal individuality (or 'mineness') of experience. However while some phenomenologists want to confine this mineness to the way that experience occurs, I argue that this mineness points to a more substantial knowing of ourselves as *bodily* subjects or selves.

I claim that this human centered starting point becomes relevant or critical when we try to construct - or claim to have knowledge of - universal metaphysical theories such as scientism, which purports to be able to give a complete physicalist scientific picture of the world that sees nothing special in our individual first person perspectives. I argue that purely third personal objective scientific metaphysical theories have a blind spot in accounting for consciousness, which is what makes possible any theory whatsoever. Also since the twentieth century the results of physical science itself, notably quantum physics, seem, on some interpretations, to point to a reality that is not independent of individual observers.

To alleviate these problems, I argue for a re-ordering of the priority of self and science. I reject strong forms of scientific realism and scientism in favour of a view that understands science as just one form (albeit an important one) of our human exploration of our world. Our knowing-by-being of ourselves is existentially and epistemically prior to any third-person knowledge and reflects our human cosmic significance as sources of explanatory knowledge and judgment. However, because our experience of ourselves is from within a wider world we should not expect to have a complete ontological picture of everything.

Our limited ontological knowledge leaves us in a situation of subject and object duality but this is not a matter of dividing the world; rather it is a matter of recognizing our own situation. This can be seen as limiting, in that we are not in a position to know everything: but a proper understanding of ourselves liberates us from universalist doctrines and allows us to know ourselves as a source of nature's self understanding from within.

INTRODUCTION

(i) SELF-REFERENTIALITY

According to Maurice Merleau-Ponty (1964, 27), philosophy is "the set of questions in which the one who questions is himself implicated in the question."

This suggests that self-reference should be a central topic in philosophy. However, it has mainly been discussed in terms of language and the paradoxes that can arise when a statement refers to itself or its own referent. As Thomas Bolander (2015) writes in the Stanford Encyclopedia of Philosophy entry on self-reference – "The philosophical interest in self-reference is to a large extent centered around the paradoxes." But if Merleau-Ponty is right, a more significant form of self-reference is what distinguishes philosophy from science. He argues that unlike science, which can "soar above its object" and assume the correlation between knowledge and being, philosophy involves a self-referential "radical examination of our belongingness to the world before all science".

In this dissertation I explore the priority of our individual being (being me) and our knowing our self (knowing-by-being), before any inquiry whatsoever. I adapt Ludwig Wittgenstein's phrase "the world as I found it" to express the starting point of any such inquiry. This coincides with the phenomenological idea that philosophy should start from a description of the way we find ourselves living. In the *Tractatus Logico-Philosophicus* (1922), Wittgenstein imagines writing a book called 'The World as I found it". He argues that the one thing that could not be mentioned in this book is the subject or self, because the subject does not belong to the world; rather it is the limit of the world (Wittgenstein, 1922, 5.631 & 2). My dissertation is not about Wittgenstein's ideas but I point to a similar distinction between the scientific picture of the world, which, on some versions, doesn't have any place for the subject that produces it. However, I differ from Wittgenstein's analysis in that I argue that the self or 'I' is a central feature of "the world as I found it": "I" in fact features centrally in Wittgenstein's phrase. I argue that

self-reference, where the 'self' is a person or agent, is central to any philosophical or scientific inquiry.

In this introduction, I examine some different kinds of self-reference, starting with some purely semantic examples.

A simple sentence such as -

- "This sentence contains five words" -

demonstrates an internal coherence between the meaning of the sentence and its linguistic expression. Understanding the self-reference requires an attention shift from one object of consciousness (what is meant) to another object of consciousness (the written words).

The liar paradox exploits this type of self-reference to produce a sentence such as -

- "This sentence is not true".

This appears to be paradoxical because if we assume the sentence to be true, then what it states must be the case, that is, it cannot be true. If, on the other hand, we assume it not to be true, then what it states is actually the case, and thus it must be true. In either case we are led to a contradiction.

This type of *semantic self-referentiality*, while initially puzzling, can possibly be circumvented, by introducing some kind of hierarchy to stratify the universe of discourse (sentences, meanings) into levels. By making a stratification in which a sentence may only contain or refer to objects at a lower level, circularity disappears. In any case language is a human invention, which we use for practical purposes, and maybe we should not expect it to be logically complete, at least in everyday settings.

However we might expect formal or logical systems to be complete. But Kurt Gödel (1962) identified certain sentences that have become known as "Gödel sentences" that exhibit a kind of *logical self-referentiality*. These sentences are superficially similar to the liar paradox sentences, but exemplify a possible problem for certain types of formal or logical systems such as mathematics. His idea was to use mathematical reasoning in exploring mathematical reasoning itself. This lead to his identification of his Incompleteness Theorem, which Douglas Hofstadter (1979, 17) paraphrases as –

- "All consistent axiomatic formulations of number theory include undecidable propositions".

In the case of Alfred North Whitehead and Bertrand Russell's *Principa Mathematica* (1910, 1912, 1913) such a proposition might state that –

- "This formula is unprovable by the rules of *Principa Mathematica".*

So there is a statement in *Principa Mathematica's* system, that if true, is unprovable, and if provable, shows that the system proves falsehoods. This, then, is not an apparently contradictory statement as the liar paradox is; rather the sentence is claimed to be a *true* statement about the formal system but the truth of the statement cannot be proved within the system itself (although we can know its truth). Gödel concludes that consistent formal systems of this sort are incomplete. This conclusion has been used in various ways – for example - by John Lucas (1961, 113) to attempt to show that "no machine can be a complete or adequate model of the mind, that minds are essentially different from machines", and by Roger Penrose (1989, 540) to argue that mathematical "insight" is nonalgorithmic.

These responses latch onto the incompleteness – but the theorem can also be read as showing that no formal system can be *both* consistent and complete. So if you can allow some inconsistency you can have a complete or at least functioning system. This might be more relevant to human minds or human intelligence. So humans can be intelligent in spite of often making mistakes and holding contradictory beliefs. As Alan Turing (2004/1948, 411) pointed out, not making mistakes "is not a requirement for intelligence".

A more significant contradictory self-reference can occur when one of the referents is a person or an aspect of a person. It is this type of self-reference that I am more interested in.

The liar paradox is also known as Epimenides paradox. In the 6th. century BC, Epimenides is reported to have written -

- "The Cretans, always liars …" or "All Cretans are liars" (Harris, 1906).

In this case the sentence is in itself not semantically self-referentially contradictory. Rather it is the fact that the speaker or writer himself was from Knossos in Crete that makes the statement seem contradictory. It

becomes self-referential (and therefore contradictory) because of something external to the content of the sentence itself.

A more recent case of potential inconsistent speaker (or *epistemic*) self-reference can be found in Norman Malcolm's Conceivability of Mechanism argument (1968). According to Malcolm mechanism assumes that there is a neurophysiological theory that is adequate to explain and predict all movements of human bodies except those caused by outside forces. The human body is assumed to be as complete a causal system, as is a car's engine. On the other hand, our daily discourse is filled with explanations of behavior in terms of the agent's purposes or intentions. Malcolm argues that there is a logical difference between neurophysiological explanations and purposive explanations. Neurophysiological explanation premises express contingent correlations between neurological processes and behavior. Purposive explanation premises express a priori connections between intentions (purpose, desires, goals) and behavior. Thus, Malcolm argues that mechanism, as a universal theory, is inconceivable not because it is self-contradictory (it is not), but rather because "the absurdity lies in the human act of asserting the doctrine. The occurrence of the act of assertion is inconsistent with the content of the assertion" (Malcolm, 1968, pp.67-68). If mechanism were a true universal theory then no genuine human assertion of any theory whatsoever would be possible.

A similar kind of apparent inconsistency is identified by Lynne Rudder Baker (1987,147) who argues that Churchland's eliminative materialism is self-defeating or self-referentially incoherent and thus faces "the threat of cognitive suicide" – "if the thesis denying the common-sense conception of the mental is true, then the concepts of rational acceptability, of assertion, of cognitive error, even of truth and falsity are called into question." She concludes, "… it seems that we can neither rationally accept nor assert nor even formulate the thesis denying the common-sense conception of the mental. Indeed, if the thesis is true, it is at least problematic whether we can rationally accept or assert or even formulate any thesis at all. This seems ample reason to deny the conclusions of the argument from physicalism". Baker is pointing out that any thesis or assertion about the world is a

product of a mind, and that therefore if the thesis makes claims about minds these must be at least consistent with the possibility of this occurring.

These arguments suggest that universal metaphysical theories such as mechanism or physicalism ought to be self-referentially coherent. If such theories are at least human assertions about how the world is, they must allow for genuine human assertion. This constraint on universal metaphysical theorizing is a central theme of this dissertation. As William Hasker (1999, 84) comments on physicalistic determinism –

> "It isn't easy to make out a rational case for accepting a theory which, if correct, would imply that one can't make out a rational case for anything at all!"

This type of self-refuting or *peritrope* argument can be traced back to the Greek philosopher Epicurus (341-270 BC) who noted that -

> "He who says that all things happen of necessity can hardly find fault with one who denies that all happens by necessity; for on his own theory this very argument is voiced by necessity "(Epicurus 1964: XL).

Robert Lockie has recently proposed a version of this argument in terms epistemic justification in relation to free will –

> " … we could not be epistemically justified in undermining a strong notion of free will, as a strong notion of free will would be required for any such process of undermining to be itself epistemically justified" (Lockie, 2018, 5).

More sophisticated physicalist theories such as non-reductive physicalism and emergence claim to address this self-refutation problem. Emergentist theories attempt to explain how features such as mental states and consciousness could arise out of physical processes. However, there is an inevitable tension in such theories in trying to reconcile our experience as conscious subjects with a physical account of reality. If the emergence is too weak it may not explain the emergent features adequately. If it is too strong the emergent features may not be sufficiently grounded in the physical. It is not clear that a coherent account has yet been given. I will return to this issue in chapter 3.

Another type of self-reference – what Shaun Gallagher (2012, 133) calls *experiential self- reference* – focuses on the use of the first person pronoun ("I") in a self–referring way. William James (1890/1950, I, 225) observed that all experience is "personal" – it is always someone's experience. He said that "the universal fact is not 'feelings and thoughts exist', but 'I think' and I feel'." Ludwig Wittgenstein (1953) pointed out that when we use the first person pronoun as an experiencing subject we are certain as to who is experiencing. Sidney Shoemaker (1984) codified this insight as IEM – immunity to error through misidentification relative to the first-person pronoun. This suggests that when a speaker uses the first-person pronoun ("I") to refer to herself *as subject,* she cannot make a mistake about the person to whom she is referring.

If this is not simply a grammatical principle but rather it depends on the kind of experiential access that I have to myself, it suggests a further type of self-reference that might be called *existential* or *radical self-reference*. As Shaun Gallagher puts it IEM is a claim "not about my *awareness* of myself as perspectival origin – it's about *being* the perspectival origin of my awareness" (Gallagher, 2012,209). It is claimed that our own existence and the way we exist is immediately present to us in our actions and thoughts.

John Perry (1979) explains how this kind of radical self-reference characterizes our 'I' beliefs. He tells a story of how he once followed a trail of sugar around a supermarket aisle to try and find the shopper who was making a mess. Eventually he realized that "I was the shopper that was making the mess". Reflecting on the change in his belief from 'someone' to 'I', he notes that the word 'I' is ineliminable in explaining his behaviour – "When we replace it with other designations of me, we no longer have an explanation of my behaviour and so, it seems, no longer an attribution of the same belief. It seems to be an *essential* indexical." To explain this he draws a sharp distinction between objects of belief and belief states. So anyone at any time can have access to any proposition - but not in any way. "Anyone can believe of John Perry that he is making a mess. And anyone can be in the belief state classified by the sentence 'I am making a mess'. But only I can have that belief by being in that state." I take Perry to be claiming that

'belief by being in [a] state' is a form of radical self-reference that is crucial for the explanation of action, and also indeed, the motivation of action.

This radical self-reference applies not just to beliefs but also to knowing. As Galen Strawson (2009, 286) claims – "We're acquainted with reality *as it is in itself,* in certain respects, in having experience as we do: the having is the knowing." Strawson is arguing that in the case of experience itself, how things appear or seem is how they really are - and we are directly acquainted with this just in having experiences. I express this as 'knowing-by-being that which is known'. As Michel Bitbol and Pier-Luigi Luisi (2011) explain it -

> "… when we raise a question about consciousness, we are not only implicated it *in abstracto,* timelessly, as generic human beings; we are fully implicated in it by what we *are* at this precise moment" (Bitbol and Luisi, 2011, 209).

Formulating, understanding and answering questions about consciousness are already acts of consciousness. So they say that the self-referential character of a question about consciousness is seen when attention shifts from the meaning of the question to present conscious experience as the background of this very act of attention. But this second focus of attention is not another 'object' of consciousness; rather, it is the condition for anything to be taken as an object. This is why Bitbol and Luisi argue that questions about consciousness are more than merely self-referential: they are radically self-referential. They define consciousness as "the name we give to the astounding realization of immediate existence" (Bitbol and Luisi, 2011, 211).

As René Descartes put it –

> "For if I judge that the wax exists from the fact that I see it, it would certainly follow much more clearly, from the fact that I see it, that I myself exist. For it may be the case that what I see is not really wax … but obviously it cannot be the case, while I see or while I seem to see … that I myself am nothing as long as I am thinking" (Descartes, Meditation 2).

Descartes' *Cogito ergo sum* is traditionally translated as 'I think therefore I am'. Jonathan Shear (1999, 407) points out however that what Descartes intended by cogito is what we would now call 'being conscious' rather than merely thinking, including, for example, willing, sensing, imagining, asserting, understanding and doubting. A better, modern English, translation is, therefore– 'I am conscious, therefore I exist'. Thus Descartes argued that one cannot coherently doubt one's own conscious existence, since the very act of doubting implies it. Put linguistically, in terms of self-reference, one can never truly say or think - 'I do not exist'. But this is not just linguistic because it expresses something about ourselves that we know by being what is known.

Developing this idea, Thomas Nagel (2012, 31) argues that –

> "Our own existence presents us with the fact that somehow the world generates conscious beings capable of recognizing reasons for action and belief, distinguishing some necessary truths, and evaluating the evidence for alternative hypotheses about the natural order."

This suggests to me what I call a form of *anthropic* self-reference (which I will consider further in chapter 7) – the way that we find ourselves living in a world tells us something about that world and ourselves. We know some of what Shaun Gallagher (2012, 2) terms 'primary facts'. In this sense we instantiate a radically self-referential knowing-by-being.

In this section I have outlined some different kinds of self-reference. The first two kinds – *semantic* and *logical* – give rise to interesting questions about language and formal systems but are not the focus of this dissertation. I am more interested in cases where the 'self' in self-reference is a person. These cases – *epistemic, experiential, existential* and *anthropic* – are not meant to be mutually exclusive. Rather they are different aspects of our belongingness to the world.

(ii) PREVIEW OF CHAPTERS

In Chapter 1, I argue that the central feature of the way we find ourselves living is as individual subjects producing explanations and performing actions, with others in a wider world. This knowledge of ourselves as subjects, while often non-explicit, is direct – we know it by being that which is known. The overall primary fact is individual first-person consciousness, giving us a proto-ontology of me, others, world with three interlocking modes – independence, intersubjectivity, and interdependence. We also know certain primary facts about ourselves in terms of unity, compositionality, temporality, judgment, agency, and teleology. I argue that this direct knowing of ourselves by being, is not subject to the elusiveness thesis, which argues that from the first-person perspective one's self is unavailable to one's own awareness.

In Chapter 2, I examine whether the phenomenological concept of pre-reflective self-consciousness indicates the existence of a metaphysical self. Some contemporary phenomenologists, such as Dan Zahavi and Wolfgang Fasching, minimize the basic concept of self so that it is no more than the way that experience occurs (Zahavi, 2008, 2014), or the dimension in which it occurs (Fasching, 2009). I argue however that such a minimal self, confined to the level of experience, is not adequate to capture our awareness of our selves from the way we find ourselves living. Rather, we need to look to what gives rise to experience. I agree with Aaron Henry and Evan Thompson's view, (2011), that we are prereflectively aware of ourselves as a living body. Knowing-by-being a bodily subject justifies a more substantial core concept of self than phenomenologists such as Zahavi identify.

Chapter 3 examines the physicalist scientific picture of the world, which purports to give an objective account of the world that sees nothing special in our individual first-person perspectives. This approach tends to prioritise some kind of external "objective" reality that can be reached through third personal science, over our own experience (which is itself seen as "subjective" and not always reliable). However, I argue that physics has a limited domain of enquiry and is self-referentially incomplete, in that it fails

to allow for an account of its own assertion. This suggests that physics, on its own, will not give us a metaphysically complete picture of reality. Furthermore, there is no agreed interpretation of what the currently best theory of physics (quantum physics) tells us about reality. Moreover, the persistent problem of how to account for consciousness in any universal physicalist picture of reality remains unresolved.

I end this chapter with a consideration of four possible 'escape routes' that William Seager (2012) suggests might lead us out of the current impasse. The first involves 'watchful waiting', that is giving science more time to come up with a satisfactory explanation of consciousness. This might seem prudent but with no clear plan of how to proceed, it might be time to try and find a different solution.

The next two options of 'embracing [radical] emergence' and 'favouring fundamentality' seem to be inextricably linked. Without some account of consciousness at the fundamental level, radical emergence seems like a brute fact that we must accept with 'natural piety'. On the other hand, any account in terms of fundamental consciousness entities faces a problem of explaining how these entities combine or divide to form the unified consciousness that we know by experiencing it. Because of the problems with these options Seager's fourth option considers the merits of 'modifying metaphysics', which involves a rejection of some forms of scientific realism and any purely scientistic metaphysics.

This is the theme of Chapter 5, but before that Chapter 4 considers the one science – biology – that at least in some cases, takes seriously and tries to explain the nature of selves as unities over and above their physical constituents. Biological structures such as cells and organisms exhibit an individuality and unity achieved by an immune system and structure. This leads to a consideration of hylomorphism – the claim that structure is a basic ontological and explanatory principle. While biology gives us some explanation of the origin and nature of selves, I argue that many of the most important discoveries of biology are based on or presuppose the knowledge we have of our selves as living beings.

This re-ordering of the priority of self and the science of biology is applied more widely to science in general in Chapter 5, which returns to Searger's fourth escape route – 'modifying metaphysics'. This involves a rejection of scientism – the position holding that science provides the best, perhaps only, route towards knowledge of the ultimate nature of reality. This does not undercut the explanatory power of science within its proper sphere, but rather is based on a view of science as abstraction or modeling, which casts doubt on the ability of science alone to produce an overall picture of reality. This is consistent with the phenomenological understanding of science, which sees it as an abstraction away from conscious lived experience. While some attempts to modify science to accommodate consciousness or subjectivity seem promising, I argue that knowing by being is prior to any scientific knowledge whatsoever, so that a comprehensive picture of reality, would have to be wider than any scientific picture of the world – it would have include what we know by being.

Chapter 6 considers the possibility that on the one hand our knowing of ourself is mistaken, and there are no selves or maybe just one self; or on the other hand, there might be group selves or that other living entities are also selves. I argue that the basic metaphysical self as I have presented it, is not the target of 'no-self' or 'one-self' theories but is rather a presupposition of these theories. I reject the possibility that there could be group selves of the same kind as individual selves but I do endorse a view that selves, in a minimal but significant sense, are ubiquitous in all living entities.

In Chapter 7, I use an interpretation of the weak anthropic principle to ask what properties an observer must have if the universe is successfully observed. This leads to the suggestion that our actions as observers and explainers of our world exemplify a form of existence that is significant, not just for ourselves, but also for the universe as a whole. Our capacity for judgment allows us create and use explanatory knowledge, which is central to the development of science and other forms of knowledge: and our ability to know ourselves as the source of this knowledge allows for the possibility that we might use this knowledge wisely.

In Chapter 8 I return to our first-person experience of ourselves as individual conscious subjects – 'being me'. Because our knowing by being of ourselves as conscious subjects is there not on the basis of experiencing, but in the act of experiencing itself, I argue that this experience is epistemically and existentially prior to any third-person knowledge. But I argue that our experience of ourselves as knowers from within a wider world means that we should not expect to have a complete ontological picture of everything. Our limited ontological knowledge leaves us in a situation of subject and object duality but this is not a matter of dividing the world; rather it is a matter of recognizing our own situation. This can be seen as limiting, in that we are not in a position to know everything: but a proper understanding of ourselves liberates us from universalist doctrines and allows us to see ourselves as a source of nature's self understanding from within.

CHAPTER 1. KNOWING-BY-BEING & PRIMARY FACTS

1.1 In the beginning was the deed.
1.2 Knowing-by being.
1.3 Primary Facts.
    1.3.1 Consciousness
    1.3.2 Independence
    1.3.3 Intersubjectivity
    1.3.4 Interdependence
    1.3.5 Unity
    1.3.6 Compositionality
    1.3.7 Temporality
    1.3.8 Judgment
    1.3.9 Agency
    1.3.10 Teleology
1.4 Elusiveness v. Direct Awareness of Self.

1.1 IN THE BEGINNING WAS THE DEED

In *On Certainty*, Ludwig Wittgenstein (1969, 396) quotes Goethe's line from Faust with approval - '*Im Anfang war die Tat*' - 'In the beginning was the deed'. Specifically, in relation to language games, Wittgenstein puts 'our acting' rather than any proposition as the grounds of any enquiry –

> "Giving grounds, … justifying the evidence, comes to an end; but the end is not certain propositions striking us as immediately true, i.e. it is not a kind of *seeing* in our part; it is our *acting* which lies at the bottom of the language game" (Wittgenstein, 1969, 204).

This suggests to me that the foundation or starting point of all enquiry is not any assumption or proposition that we might discover, but rather, the very action that we find ourselves performing as subjects producing explanations and acting in an on-going world. This experience of ourselves as knowers and actors seems undeniable because any consideration of the

possibility that this is a radical illusion would have to involve an act of knowing which would contradict the illusion.

I find myself as a small part of an on-going world. Whatever direct knowledge of the wider world I have is my knowledge from my unique place in the world. And any other knowledge I can get will be based on other people's reports of their knowledge from their perspective. And whatever I do, under pressure of circumstance or voluntarily will be my action – I will have to do something to achieve it, and I will reap the benefit or suffer the consequences of my action. So, I find myself situated in this on going world, thinking and acting from my unique personal perspective. The most pervasive feature of the world as I come to know it is that all experience is personal. Edmund Husserl pointed out that experiences do not just lie about like stones and chairs, equally accessible to everyone – "experiences are not in a nowhere-land" but "in me" (Husserl, 1956, 166). As William James (1890/1950, I, 225) observed - "the universal fact is not 'feelings and thoughts exist', but 'I think' and I feel'."

But who am I, who thinks and feels?

There are different ways that I might go about finding out about myself. So for example science tells me that I am a particular type of animal - a human being, and gives me lots of detailed information such as sociological, psychological, biological, chemical, physical etc. Culture tells me about my family, my society, my country, and my planet. But neither science nor culture can give me individual knowledge about myself. By reflectively thinking about myself (introspection) in the way I think about other things in the world, I will learn certain facts about myself, but not what I am. This dissertation aims to draw attention to a more basic (pre-reflective) sense of what we are by being aware of the way we find ourselves living. From the way I find myself living I have a sense of myself as the subject of experiences and the agent of actions, and also as a unique individual among others. I call this sense of myself knowing-by-being.

1.2 KNOWING-BY-BEING

One way to approach knowing-by-being is in relation to Bertrand Russell's distinction between 'knowledge by description' and 'knowledge by acquaintance'. In his *Problems of Philosophy,* Russell (1912) introduces this distinction to clarify the relationship between sense-data and the objects that cause them. He gives "the name 'sense data' to the things that are immediately known in sensation: such things as colours, sounds, smells, hardnesses, roughnesses, and so on" (ibid. 12). He says that we have *acquaintance* with sense-data because they are that of which "we are directly aware, without the intermediary of any process of inference or any knowledge of truths" (ibid. 46). Knowledge of objects, on the other hand, is not direct knowledge because "it is obtained through acquaintance with the sense-data that make up the appearance of the [object]" (ibid. 46-47). He calls knowledge of objects 'knowledge by description'. In the case of a table he says that "(m)y knowledge of the table is of the kind which we shall call 'knowledge by description'. The table is 'the physical object which causes such-and-such sense-data'. This *describes* the table by means of the sense-data" (ibid. 46-47).

The two modes of knowing are distinguished by their certainty – "it is possible, without absurdity, to doubt whether there is a table at all, whereas it is not possible to doubt the sense-data" (ibid. 46-47). Russell argued that we have acquaintance not just of sense-data through sensation, but also of our thoughts, feelings and desires through introspection, and of past data through memory.

For Russell then, some knowledge is more direct and therefore more certain than other knowledge. Russell further considers the possibility that we are acquainted with our "bare selves as opposed to particular thoughts and feelings". On the one hand he agrees with David Hume (1739/1975, 252) that such a feat is not possible - "When we try to look into ourselves we always seem to come upon some particular thought or feeling, and not upon the 'I' which has the thought or feeling" (Russell, 1914, part III).

On the other hand, he argues that we know the truth 'I am acquainted with this sense-data', and, "(i)t is hard to see how we could know this truth, or even understand what is meant by it, unless we were acquainted with something which we call 'I'." So he says "in some sense it would seem we must be acquainted with our Selves as opposed to our particular experiences". Russell's conclusion on this is uncharacteristically non-committal - "the question is difficult, and complicated arguments can be adduced on either side. Hence, although acquaintance with ourselves seems *probably* to occur, it is not wise to assert that it undoubtedly does occur." And, he later denies that we have acquaintance with the self (Russell 1914: part III).

But maybe there is another way that we know the self. My suggestion is that there is a further kind of knowing, namely knowing–by-being, which is both more direct and more certain even than knowledge by acquaintance, and therefore more fundamental. Our awareness of ourselves as the subject who knows is present directly in every act of knowing. As Samuel Alexander (1924) explains it – "In knowing the object I know myself, not in the sense that I contemplate myself, for I do not do so, but in the sense that I live through this experience of myself" (Alexander, 1924, I.xiv). And living through the experience of myself as subject is what Louis Saas (1998) identifies as the most fundamental sense of selfhood –

> "The most fundamental sense of selfhood involves the experience of self not as an object of awareness but, in some crucial respects, as an unseen point of origin for action, experience, and thought … What William James called … the 'central nucleus of the Self' is not, in fact, experienced as an entity in the focus of our awareness, but, rather, as a kind of medium of awareness, source of activity, or general directedness towards the world" (Saas, 1998, 562).

In this special case of knowing then, we are both what is known and what knows. Because of this identity our knowledge is both direct and certain, in a way stronger and more fundamental than that suggested by knowledge by acquaintance. As Martine Nida-Rümelin (2012) puts it, we should accept a thesis of revelation with respect to what it is to be an experiencing subject –

> "… in a non-conceptual and implicit manner we are aware of what it is to be an experiencing subject *by being* an experiencing subject; we thus have implicit knowledge of what it *is* to be an experiencing subject given that we *are* subjects of experience; the nature of what it is to be an experiencing subject is revealed to us by being a subject of experience" (Nida-Rümelin, 2012, 18).

I make a clear distinction between knowledge by acquaintance of aspects of ourselves, and knowing-by-being. Galen Strawson (2011, 287-8) seems to take knowing by being as just a special case of acquaintance knowledge – he calls it "knowledge by direct acquaintance" (ibid. 296). I argue however that our knowing-by-being of ourselves as subjects is distinct and different in quality from the knowledge we might have by acquaintance with our sensations, thoughts, or memories. In Russell's terms we have descriptive knowledge of ourselves through inferences from what we know more directly. We also have acquaintance knowledge of aspects of ourselves through sensation, introspection and memory. I claim that the knowledge that we have of ourselves as the subject that has the sensation, introspection and memory is a further kind of knowledge that is more basic or fundamental.

In *Selves,* Strawson (2009, 286) does seem to accept that the term 'acquaintance' may be misleading – "The word 'acquaintance' may be thought to introduce too much of an idea of relation. This reveals too narrow a notion of what a relation is." Maybe so, but identity seems to be qualitatively different from any other relation whatsoever. I suggest that the identity that we experience as subjects is better expressed by 'knowing-by-being'– for as Strawson himself says - "the having is the knowing" (Strawson, 2011, 286). This would leave the term 'knowledge by acquaintance' to refer to situations (of sensation, introspection and memory) in which there is some distance in the relation between the knower and the known, whereas "knowing-by-being" refers to direct knowledge we have of ourselves as the subject of any possible experience.

Through knowing-by-being we have a subjective sense of our existence - a sense of an "I" that underlies all our desires, activities and physical

characteristics. Arthur J. Deikman (1996, 350) argues that this "I" is conceptually prior to all conscious experience - "The core 'I' of subjectivity is different from any content because it turns out to be that which witnesses – not that which is observed. The 'I' can be experienced, but it cannot be 'seen'. 'I' is the observer, the experiencer, prior to all conscious content." Deikman claims that "I am sure I exist because my core 'I' is awareness itself, my ground of being." For this internal observer, even bodily feelings are observed.

Deikman claims that we know this internal observer not by observing it but *by being* it. "At the core, we are awareness and therefore do not need to imagine, observe, or perceive it." So he claims "(k)nowing by being that which is known is ontologically different from perceptual knowledge".

Thus, we don't know ourselves only by perceiving them, describing them, or becoming acquainted with them; we also know ourselves by *being* them - I know myself by *being me.*

1.3 PRIMARY FACTS

Knowing-by-being expresses the realization that we know something about ourselves and the world, just from the way we pre-reflectively find ourselves existing. We find ourselves as conscious subjects producing explanations and acting in an on-going world: but also as a unique individual among others. From this realization we can identify several factors that constitute our human way of being. Shaun Gallagher (2012, 2) calls these 'primary facts'.

These facts are 'primary' in the sense of being 'conceptually more basic' or 'prior', rather than necessarily 'primary' in any developmental sense. As Michel Bitbol and Pier-Luigi Luisi (2011, 211) put it, they "just state the elementary features of what we are living". This is what Edmond Husserl (1913/1931) called a phenomenological description: a plain statement of what is immediately experienced, irrespective of any interpretation of the contents of experience in naturalistic terms. Husserl called the collection of

situations - both natural and cultural - in which we find ourselves involved, our *lifeworld.* As Shaun Gallagher (2012) explains it, our lifeworld "is the world as we live it, not just the world as it opens up in front of us as perceiving subjects, but the world which is at the same time something *already there* operating as the meaningful background for all our actions and interactions". So the lifeworld is the world we take for granted rather than the world as we study it through science, or represent it through art. Thus Gallagher says that the lifeworld "is not the world that we take as object, as something distinct from ourselves, but is rather a specification of our existence" (Gallagher, 2012, 2).

In Wittgenstein's terms these primary facts make up our world-picture ("The world as I found it") that is not learned by experience but rather is the inherited background against which we distinguish between true and false (Wittgenstein, 1969, 94). Children do not learn the facts that make up this world-picture from a book; rather, "they as it were swallow them down with what they do learn" (Kenny, 2010, 882) -

> "When a child learns language it learns at the same time what is to be investigated and what not … Children do not learn that books exist, that armchairs exist, etc. etc., - they learn to fetch books, sit in armchairs etc. etc." (Wittgenstein, 1969, 472, 476).

And this inherited background is not just transmitted culturally, but also, more basically, we are 'prewired', in a process called neurognosis, to know ourselves, and the world around us in a certain way. As Laughlin and Throop (2009) describe it –

> "The fundamental structures are genetically organized very early in the womb and the first months of life. We are born knowing the world from a very human standpoint, our world of experience being (in a sense) "already there" due to the intricate organization of neural structures and sensory systems already laid down at or before birth."

Neurognosis lays the intuitive foundations for cultural cosmologies and is responsible for certain ubiquitous elements of the worldviews. However, David Deutsch argues that although we are born with a certain worldview this need not constrain us because we also have an ability to improve or

expand it. So while "(w)e do not begin with 'white paper' at birth, but with inborn expectations and intentions", we also have "an innate ability to improve upon them using thought and experience" (Deutsch, 2012, 4).

1.3.1 CONSCIOUSNESS

Starting from "the world as I found it" the overall primary fact is that there is consciousness in the simple sense of 'phenomenal consciousness' or experience. We experience things.

In this dissertation I use consciousness in this first-personal experiential sense rather than any intra-mental (or informational processing) sense such as 'access consciousness', which understands a state's being conscious as a matter of its availability to interact with other states, and of the access that one has to its contents. In section 6.23 I present an argument that suggests that phenomenal consciousness is the core concept of consciousness and this is presupposed by other concepts such as access consciousness.

Consciousness, in this core phenomenal sense then, is not an object in the world. Rather consciousness is our openness to (or the disclosure of) the world. Thus, as Guilio Tononi (2015) puts it consciousness exists intrinsically – "my experience exists from its own intrinsic perspective, independent of external observers", and "that my experience here and now exists (it is real) is the only fact I can be sure of immediately and absolutely".

Consciousness is the fact that there is a 'world' there for me in a certain way. As Shaun Gallagher points out - "our only access to the way things are in reality (and even to the very idea that there *are* things in reality; even to the very idea that there is something we call 'reality') is *via* consciousness (Gallagher, 2012, 8).

Thomas Nagel (1974) argues that a thing is conscious just in case there's something that it's like to be it; if it has an 'inner life' of some kind. And the primary way we find ourselves living is as having an inner experience of an external world in a certain way. Also, this primary conscious experience is not an experience of the world in general. It is always from a first-person

perspective. The 'I', or one who is conscious, is present in every conscious experience. As Sartre put it self-consciousness is "the only mode of existence which is possible for a consciousness of something" (Sartre, 1942/1969, xxx). For Sartre then, an experience does not simply exist; rather, it exists in such a way that it is implicitly self-given – 'for itself' (*pour-soi*). And this self-giveness is not something added to the experience; rather, the very mode of being of consciousness is to be self-conscious. Sartre then argues that consciousness, because of its ubiquitous pre-reflective self-consciousness, must possess a basic dimension of selfhood, that he calls *ipseity* (from *ipse* – the Latin for self).

This primary datum of consciousness can be spelled out in terms of three proto-ontological categories – self, other, world. These are proto-ontological in that they describe the situation we find ourselves living in prior to any possible theory – ontological, scientific or otherwise. My own knowledge of anything is from my own first person perspective; but I am always aware of others with their own perspectives; and we live in a larger world that affects us, and is affected by us. As Martine Nida-Rümelin (forthcoming) puts it –

> "The notion of an experiencing subject is deeply incorporated in the way we experience the world and in the way we think about other human and non-human animals that we take to be conscious. We have an intuitive and, in a sense, clear understanding of the difference between a mere thing, without any 'inner life' or 'perspective' and an individual, an elephant, for example, for whom it is like something to live through its individual existence. We have a substantial positive understanding of the difference between an experiencing subject and a complex information processing thing without 'the light of consciousness'".

Being experiencing subjects ourselves, the primary distinction we make in the world as it appears to us, is between other conscious subjects and "mere things".

The three categories (self, others, world) have three interlocking modes of existence – independence, intersubjectivity, and interdependence.

1.3.2 INDEPENDENCE

One primary fact is that we find ourselves existing as somewhat independent individuals separated in some way from others, and the rest of the world. As Guilio Tononi (2015) puts it, experience is exclusive, in that it has "definite borders, temporal and spatial grain".

Dan Zahavi points out that an essential aspect of experience is that "we have a distinctly different acquaintance with our own experiential life than with the experiential life of others (and vice versa), and … this difference obtains, not only when we introspect or reflect, but already in the very having of the experience" (Zahavi, 2020, 8). Our experience is presented to us in terms of a difference between our psychophysical body and what ever is external to us, including other people.

Edmond Husserl (1960, 97) identifies different ways in which we experience this independence. In a subject's experience one animate organism is '*uniquely* singled out' due to the fact that firstly, the subject can ascribe hyletic experiences (e.g., tactile sensations or sensations of warmth or coldness) only to the subject's own body; secondly, the subject has the ability to '*rule and govern'* this body '*immediately'*; and thirdly, every bodily movement is accompanied by kinesthetic sensations that belong only to the organism undertaking the movement. All such experiences help to distinguish the subject's body as separate from everything else.

We also become aware of ourselves as inner, psychological subjects – what Galen Strawson terms the 'thin subject'. The thin subject is "the subject considered specifically as something 'inner', something mental, the 'self', if you like, the inner 'locus' of consciousness considered as such" (Strawson, 2011, 286).

Whether we consider ourselves as thin subjects or bodily subjects (or both), we always are aware of ourselves as somewhat independent from everything else, such that we have our own perspective on the world. Our own awareness of ourselves having private thoughts and a unique point of view on the world is the source of our sense of independence. And as Di Paolo and Thompson (2014, 2) point out, we are not just independent

individuals, but rather, we (and all living systems) are self-individuating; we generate and maintain ourselves through constant structural and functional change, and actively establish and police the distinction between ourselves and our environment.

I will return to this idea that we are self-individuating individuals later in this dissertation, particularly in Chapter 4.

1.3.3 INTERSUBJECTIVITY

One further primary fact, that Shaun Gallagher identifies, is intersubjectivity – the fact that we find ourselves living in a world already populated with other subjects like us. Thus he writes -

> "Before we have a chance to think about this – before we have a chance to theorize, philosophize or 'phenomenologize', before we have a chance to stake out a position of any sort – our capabilities for so doing have already been shaped by other humans who have been with us even before our birth, who have been talking to us even before we could respond, and who have been interacting with us even before we could act on our own"(Gallagher, 2012, 3).

Thus, intersubjectivity is a primary feature of our world that is normally taken for granted. It is not something that is just inferred or derived from experience; rather it is part of the fabric of the world that we find ourselves living in. This normally unremarked upon close social interaction is exposed by the unnatural 'social distancing' that we are told to practice in an abnormal pandemic crisis situation in which we are trying to slow the spread of a virus.

Pre-reflectively the existence of 'other minds' is not a 'problem' that we must solve; rather our own knowing of ourselves gives us empathetic insight to understand others as beings like ourselves. Thus, Dan Zahavi (2014) claims that there is a basic sense in which we perceive the state of the other. We do not need to simulate the other's emotion in ourself, or ascribe an emotion to the other – rather we directly perceive joy or sadness

in the face of the other. But, Zahavi makes clear that although empathy is the experience of the embodied mind of the other, this does not eliminate the difference between self-experience and other-experience; rather it takes the asymmetry to be a necessary and persisting existential fact. As he puts it – "Whereas the experience of empathizing is first-personally given, the empathized experience is not given first-personally to the empathizer"; thus, "To empathically understand that your friend loves his wife is quite different from loving his wife yourself" (Zahavi, 2014, 150-151).

Similarly, John Searle argues that we have a "biologically primitive sense of the other person as a candidate for shared intentionality" and this is "a necessary condition of all collective behavior and hence all conversation" (Searle, 1990, 105). He argues that in order to have or act on collective intentions one must suppose that "others are agents like yourself, that they have a similar awareness of you as an agent like themselves, and that these awarenesses coalesce into a sense of us as possible or actual collective agents" (Searle, 1990, 414).

The relationship between our independence and intersubjectivity can be understood in terms of Hannah Ardendt's concept of plurality. According to Ardendt (1958), the cornerstone of the human condition is plurality, which means that "men, not Man, live on the earth and inhabit the world". But plurality entails two aspects: equality and difference – we are all humans but everyone is exceptional in her or his uniqueness (Ardendt, 1958, 7-8).

1.3.4 INTERDEPENDENCE

Although we (individually and collectively) are in some ways separated from the world, we are also in some sense part of, and interdependent with, the world. We are in, and of, the world. Martin Heidegger (1973, 78) expresses this fundamental interdependence by saying that we are "beings-in-the-world". The hyphens indicate that "being", "in" and "world" are not ontologically separable, but form one irreducible and unified structure, what he terms "a *unitary* phenomenon". As Brian Cantwell Smith puts it –

> "Everything we believe, everything we take in, everything we represent and are committed to, must be something we can understand as being in a single world – the world that both we and it inhabit" (Smith, 2019, 97).

The way in which we exist as being autonomous from, yet interdependent with, the world is explained by Evan Thompson (2007) in terms of autopoiesis. Autopoiesis (self-making or self-creating) is a term introduced by Chilean biologists Humberto Maturana and Francisco Varela (1973) to define the self-maintaining chemistry of living cells. This is achieved by means of an active semipermeable boundary or membrane, which regulates interaction with the outside environment. Thompson (2007) identifies the emergence of autopoiesis as the biological forerunner of the human self. He describes the animate form of our living body as the place of intersection for numerous emergent patterns of selfhood and coupling - cellular, somatic, sensorimotor, and neurocognitive. These patterns derive from distributed networks with operational closure.

Living beings affirm their own identities by differentiating themselves from their surroundings and thus demand to be seen from an autonomy perspective. As Hans Jonas (1966, 82-83) argues – "The introduction of the term 'self', unavoidable in any description of the most elementary instance of life, indicates the emergence, with life as such, of internal identity – and so, as one with emergence, its self-isolation too from all the rest of reality." I will return to the continuity between biological cells and human selves in chapter 6.

But Thompson (2007, 150) points out that self-isolation does not imply outright independence. The organism is in and of the world, and its identity has to be enacted in the process of living, which is to say in the assimilation of and accommodation to the world. Thus, autonomy, far from being exempt from the causes and conditions of the world, is an achievement dependent on those very causes and conditions. The autonomy of the organism is achieved by operational closure but thermodynamic openness.

According to Hans Jonas (1966) the organic form stands in a dialectical relation of *needful freedom* to matter. The sheer fact of metabolism brings

with it a dialectical relation of freedom and necessity peculiar to living beings. According to Steven Rose (1997) every organism enacts a "lifeline" – a continuity of the individual through material change. With this freedom comes a correlative necessity; the organism has to change; stasis is impossible. The organism's condition is thus one of needful freedom. I will return to this in chapter 4.

Thus, from where we find ourselves looking outwards we know some primary facts about our world - we are located with others in a world. From the way we find ourselves selves living we also know certain further primary facts about ourselves. The first three – unity, compositionality and temporality – identify dimensions of our existence; the second three – judgment, agency and teleology - identify capacities that we possess.

1.3.5 UNITY

Consciousness, as we experience it, is *unified*. As Guilio Tononi (2015) puts it – "each experience is *irreducible* to non-interdependent, disjoint subsets of phenomenal distinctions. Thus, I experience a whole visual scene, not the left side of the visual field independent of the right side (and vice versa)".

Also, thinking and acting are singular events that require a unified subject. As Peter van Inwagen (1990, 5) puts it – "mental predicates (like 'is in pain' or 'is thinking of Vienna') require a subject. In the case of any particular episode of thought or sensation, there must be a thing, one thing, that is doing the thinking and feeling." He argues that thinking cannot be regarded as a cooperative activity. Rather there must be a 'one', a unified individual thing that thinks. In chapter 6 I consider, but reject, the idea that there could be a group mind that could think or act as we do.

Van Inwagen argues that "… things cannot work together to think or, at least, things can work together to think only in the sense that they can compose, in the strict and mereological understanding of the word, an object

that thinks" (van Inwagen, 1990, 119). So "… the fact that I am a thinking being shows that there is at least one composite material object."

Our pre-reflective experience of ourselves as thinkers and doers is precisely as a unified 'I'. Before we do some action we may be 'in two minds' about whether we should do it or not, but at the moment of action one choice has to be made and once acted upon it becomes our action. Robert Kane (2005, 411) calls such actions "will-setting" because "the wills of agents (their motives and purposes) are not already 'set one way' *before* they act, … but rather the agents set their wills one way or the other in the performance of the actions themselves". By making decisions and performing actions we become aware of ourselves as unified subjects and agents. As Kane puts it – the agent makes "the reasons for preferring one option prevail at the moment of choice *by* choosing or deciding" (Kane, 2005, 412). Thus, choices and decisions are enacted by the agent, in virtue of this 'primary' unity.

1.3.6 COMPOSITIONALITY

Although our pre-reflective experience of ourselves as thinkers and doers is as a unified 'I', within this we can identify different aspects of ourselves. As Guilio Tononi (2015) puts it – "experience is structured, consisting of multiple aspects in various combinations".

So for example, sometimes we think rationally but often we rely on intuition or a combination of the two. We might think of ourselves as primarily conscious beings but we are also aware that often how we actually behave is the result of unconscious processes. We experience ourselves as being in some ways like other living creatures but in some ways very different.

According to Jonathan Evans' Dual Process Theory (Evans, 2003) we have two rather different cognitive styles – reason and intuition. Reason is serial, language dependent, slow, explicit, can be hypothetical, and uses working memory. Intuition, on the other hand, is completely automatic, modular

system (e.g. facial recognition), fast, unconscious, implicit, can run in parallel (because there is no need for attention or the use of working memory). In practice, a lot of performances are hybrid, in which we are making use of the two different styles of cognition simultaneously and to varying degrees.

According to John Bargh (2017, 12), Freud presented the unconscious mind as "a seething cauldron of maladaptive complexes bent on causing us trouble and grief, which could only be overcome through the intervention of our conscious mind (with the help of a good psychiatrist, of course)". On this view what is conscious is *good,* and what is not conscious is *bad*. More recent understandings of the unconscious however recognize that the unconscious has evolved to help us survive and thrive, precisely because it is fast, and therefore enables us to react quickly (Kahneman, 2011). Our conscious mind also evolved as a kind of steering wheel to allow for additional, strategic control of the unconscious mechanisms. As Bargh puts it –

> "We have a single, unified mind that operates in both conscious and unconscious modes, always using the same set of basic machinery, fine-tuned over the course of evolutionary time. The hidden mind – the mental processes operating outside our knowledge and intention – exist to help us, though it does have an array of complex effects that we will benefit from understanding" (Bargh, 2017, 13).

Hutto and Myin (2017) describe our minds as 'duplex'. According to them the basic sorts of cognition that our brains help to make possible are fundamentally interactive, dynamic, relational, and contentless (xiv). Acquiring the capacity for cognition that involves content is a special achievement. Creatures capable of contentful cognition, on this view, will have had to master very special kinds of scaffolded practices—practices involving public norms for the use of symbols, where such norms depend for their existence on a range of customs and institutions. (11)

Lynne Rudder Baker (2007) points out that in some ways we are like other living creatures, but in other ways we are radically different.

"Only we have art, science, technology, religion; only we can fiddle with the course of evolution; only we can inquire into what kind of beings we most fundamentally are."
Thus while we are a part of nature, we are morally and ontologically different from every other kind of thing in nature. As Baker puts it – "we are in the world but not of the world." (Baker, 2007, 2)

A primary fact about the way we find ourselves living then is that we have several aspects – for example - an intuitive, unconscious, animal part and a rational, conscious, uniquely human part. However, these are aspects of the one self that thinks and acts as a unity. The adaptive, intuitive, largely unconscious aspects of ourselves are the foundations for, and a necessary part of, the rational, conscious aspects that are characteristic of us as human beings, and these are 'primary' in that sense.

1.3.7 TEMPORALITY

A further primary fact about the way we find ourselves living is that we know ourselves as temporal beings. It's not just that we experience time passing; we do, but more fundamentally we ourselves are continually changing with time. We have a past, a present and maybe a future; and our present as we experience it incorporates our past and the possibilities of our future.

As Husserl (1969) argued our lived presence has a certain width or duration. He conceived our stream of conscious experience as a continuous synthesis of what is *no-longer* ('retention'), what is *now* ('primal impression'), and what is *not-yet* ('protention'). While we have a primal impression of a object that is simultaneous with our current phrase of consciousness, this is always accompanied by retentions which retain previous phases of consciousness, and protentions which anticipate experience that is about to happen. Thus the current note in a melody or the current word in a sentence is heard as part of what has gone before and what is to come.

But it is not just a matter of our retention of our immediate past in conscious experience or even our explicit recollection of our more distant past. More significantly it is the history of our lived body that increasingly becomes the expression of our unique individual existence over the course of our life. Our bodily memory (Fuchs, 2017) preserves our life experiences as dispositions, inclinations, habits and skills, and these are available to us as a capacity to act without the need to remember earlier situations in which they were learned. Once I have learned how to type or play the piano I no longer need to consciously remember which key to press – the knowledge is in my fingers. This continual embodiment of our experience (positive or negative) integrates our entire past within our present bodily self. As Maurice Merleau-Ponty explains –

> "… I am not myself a succession of 'psychic' acts, nor for that matter a nuclear *I* who brings them together into a synthetic unity, but one single experience inseparable from itself, one single 'living cohesion', one single temporality which is engaged, from birth, in making itself progressively explicit, and in confirming that cohesion in each successive present" (Merleau-Ponty, 1962, 407).

We are temporal beings then, not just from our explicit knowledge about ourselves and our biography, but also in that our bodies continually instantiate our accumulated and sedimented history, and anticipate our future. What we are - the self we know by being it - is experienced as an on-going work in progress. Being me always involves not only being but also becoming. The significance of our temporal bodily nature will be developed in chapter 2.

1.3.8 JUDGMENT

We find ourselves producing judgments about the world, and we must assume that we have sufficient ability to do this. As Brian Cantwell Smith (2019) puts it, human epistemic achievement is not just about being embodied and embedded in this world but also involves a capacity to

"recognize it *as* world". This involves more than merely calculative rationality – what Smith terms "reckoning"; rather it involves distinguishing appearance from reality, and choosing reality; recognizing the difference between right and wrong - and choosing right; distinguishing truth, falsity, and impossibility – and choosing truth. On this view the capacity for judgment involves being existentially committed to our own existence and to the integrity of the world as world (Smith, 2019, 105). As Thomas Nagel writes - we are "conscious beings capable of recognizing reasons for action and belief, distinguishing some necessary truths, and evaluating the evidence for alternative hypotheses about the natural order" (Nagel, 2012,31).

It is not just that we are sufficiently separated from the world that we can have a perspective on it, or that we are sufficiently in touch with the world that we have information of it, but that we have a capacity for judgment about the world and thus an ability to gain explanatory knowledge of the world. And as David Deutsch argues that - "[t]he ability to create and use explanatory knowledge gives *people* a power to transform nature which is ultimately not limited by parochial factors, as all other adaptations are, but only by universal laws" (Deutsch, 2011, 56). I will return to these ideas in chapter 7.

In intentional action, as in judging, we have a form of consciousness characteristic of rational agency. As John McDowell argues – "… capacities that are conceptual, capacities that belong to their possessor's rationality, are operative not only in reflective thought and action but also at the ground-floor level at which there is absorbed coping and acting in flow …" (McDowell, 2013).

E.J.Lowe places our capacity for rational judgment alongside our existence as a datum that cannot be denied –

> "a datum that is not to be denied is that we ourselves exist and have psychological powers that are definitive of ourselves as persons. This is simply an immovable framework assumption of rational debate in this or any other area of philosophy or science, for to deny one's own

existence or rationality is automatically to exclude oneself from any such debate" (Lowe, 2008, 171).

Lowe is pointing out that the way that we find ourselves asking questions about our place in the world displays a certain capacity for rational judgment, and this capacity is therefore primary in this sense.

1.3.9 AGENCY

By acting in the world we experience ourselves as, to some extent, the cause of our actions and the authors of our lives. We experience ourselves as having the capacity to exert a causal influence on our own body and on the environment. As Samuel Alexander points out the primary and most immediate experience of causality is from our own experience of our minds-

> "You are conscious of your own power in willing in so far as you experience the continuous transition of an idea of some end into the consciousness of taking the final steps to its attainment … that experience is the experience of power or activity" (Alexander, 1920, 9).

Alexander makes clear that this does not involve a reflective comparison of your action with a notion of power or activity, which finds that the action falls under that designation. Rather, as he puts it – "It is itself the experience of exerting power" (Alexander. 1920, 9). Causal power is therefore a primary fact about the way that we experience ourselves living: we know it by doing it.

As Jacob Howhy and John Michael (2017) argue, being a self is in a fundamental sense tied to existence and to agency, rather than merely to having a certain narrative, certain character traits, certain desires, or a certain bodily configuration. As they state it –

> "if I exist I must be acting, and if I act I must exist. Put differently, the highest-level belief is that I am a cause in the world, or an agent" (Howhy and Michael, 2017, 13).

On this view, the belief that I exist is equivalent to the belief that I act to maintain myself in certain states.

> "The reasoning is that I cease to exist if I cannot maintain myself in a limited number of states, that is, if my body begins to disperse across many states. I must act to prevent dispersion: inaction leaves me open to the unfettered impact of the second law of thermodynamics" (Howhy and Michael, 2017, 12).

This linking of existence and causal power is consistent with Plato's Eleatic Stranger in the Sophist who argues that –

> "everything which possesses any power of any kind, either to produce a change in anything of any nature or to be affected even in the least degree by the slightest cause, though it be only on one occasion, has real existence. For I set up as a definition which defines being, that it is nothing else but power" (Plato, *Sophist*, 247 D, E, translated by Harold N. Fowler, Loeb).

Robert Kane (2005, 426) argues that, at least in some undetermined situations, we have the capacity to make self-forming choices which are consistent with, but not determined by, the past. In these situations we are involved in a "value experiment" whose justification lies in the future and is not fully explained by the past. Thus he argues that agents who exercise free will are both the authors and the characters in their own story at once. "By virtue of 'self-forming' judgments of the will (*arbitria voluntatis*), they are 'arbiters' in their own lives, 'making themselves' out of a past that, if they are truly free, does not limit their future pathways to one" (Kane, 2005, 426). What it is to have an experience always already contains the pre-reflective sense that one has the capacity to causally interact with the things experienced.

1.3.10 TELEOLOGY

If we can identify our explanations and actions as 'my' explanation or 'my' action, then we must assume that we can have a reason or purpose for making any explanation or doing any action, and that this reason or purpose

is uniquely 'mine'. Kane says that in a partly undetermined world we can have free will in the sense of having ultimate control over our lives – "a control related to our being to some degree the ultimate creators or originators of our purposes or ends and hence ultimate 'arbiters' of our own wills" (Kane, 2015, 433).

Di Paolo and Thompson's (2014) enactive concept of autonomy provides a naturalistic way to account for what Kant described as the intrinsic teleology or immanent purposiveness that appears to belong to living beings by virtue of being self-organizing. I will return to this in chapter 6.

I have identified six possible 'primary facts' about ourselves. The first three - unity, compositionality, temporality identify the dimensions of our existence. The second three - judgment, agency and teleology – identify capacities that we possess. This list is not claimed to be complete or that all are necessary. However, they are, I claim indicative of the kind of features about ourselves that we know-by-being, and I will use them throughout this dissertation.

In the next section I will consider an argument against knowing-by-being. By seeing how this objection can be countered, I will clarify how we have direct awareness of ourselves.

1.4 ELUSIVENESS V. DIRECT AWARENESS OF SELF

The elusiveness thesis about the self might seem to argue against the possibility of knowing-by-being. The elusiveness thesis states that from the first-person perspective one's self is particularly unavailable to one's own awareness. As David Hume observed –

> "Whenever I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never catch myself at any time without a perception and never can observe anything but the perception" (Hume, 1739/1975, 252).

This suggests that it is the nature of subjectivity that the subject itself is not an object of awareness from the subject's own point of view. It is argued that the subject's mental properties are intentional and therefore make us aware of the world instead of ourselves.

Gilbert Ryle calls this the "systematic elusiveness of 'I'" (Ryle, 1949, 186). He argues that "(t)o concern oneself about oneself in any way, theoretical or practical, is to perform a higher order act", and, "(a) higher order action cannot be the action upon which it is performed". Therefore, "(s)elf-commentary, self-ridicule, and self-admonition are logically condemned to eternal penultimacy". But "nothing that is left out of any particular commentary or admonition is privileged thereby to escape comment or admonition for ever. On the contrary it may be the target of the very next comment or rebuke." So while my past self can be exhaustively described and accounted for, my present self "perpetually slips out of any hold of it that I try to take" (ibid. 187). He uses the analogy of a finger which cannot point to itself and a missile which cannot be its own target (ibid. 188). The 'finger' and the 'missile' have a subject-like relationship to their object (what is pointed to, and, the target).

This analysis argues that it is impossible to observe one's own current experience, because this presupposes a split between subject and object while in this case the object is nothing else than the subject itself. This involves an assumption that part of the subject engages in second-order observing or monitoring of first-order mental processes. However, others have argued that introspection merely involves a modified version of those very first-order mental processes.

Thus, Franz Brentano (1874, 98) wrote that inner observation cannot be the 'true source of psychology', for observing a mental event by fully focusing one's attention towards it would just lead to its disappearance. Rather, the true source of psychological inquiry is inner *perception,* that does not require that attention be focused on some mental object, but only that, when attention is focused on some (usually external) object, it remains broad enough to notice other events such as the mental processes that

underlie the act of attending, in the same way that one can *perceive* a vibration of the telescope while *observing* a planet.

Similarly, Husserl (1931) proposed phenomenological reduction as a method to give access, not to the 'inner world', but rather to the *whole field of pure experience* before exclusive intentional focusing has narrowed down the region of our full awareness. On this view then, 'reduction' represents in fact an enlargement of the span of experience. Within the flux of lived experience, Husserl claimed there is a division between *certain* (immediate and complete) and *uncertain* (mediate and incomplete) perception. Perception of immediate lived experience is certain because the *way it appears* coincides with *the way it is,* whereas perception of spatial objects is uncertain because at each moment they present themselves through partial profiles (or 'adumbrations') whose spontaneous ontological interpretation can later be disconfirmed. For Anthony Marcel (2003), 'non-observational awareness' involves defocusing of the field of attention performed in 'inner perception'.

Therefore, I argue that when I am in a conscious mental state, an awareness that the mental state is mine is generated, even though it is not presented as an awareness of my self. The self is, as it were, pre-reflectively implicit in conscious states, but is not *presented* by them.

And, Robert Howell argues that the self-awareness implicitly presented in conscious states does not contradict the elusiveness thesis –

> "In knowing that I am thinking, I am not presented with an object, but I come to know it as a fact about myself in virtue of having the sensations. The self-awareness generated by them is propositional, therefore, but not objectual. So although my mental states – by virtue of their indubitably being mine – do generate a sort of self-awareness, it is not the sort that contradicts the elusiveness thesis" (Howell, 2010, 479).

Knowing-by-being implies that our self-awareness is direct but implicit, rather than propositional, objectual or inferential.

As Galen Strawson puts it there is a "non-thetic form of self-apprehension in which the I or subject – or just consciousness, if you wish – can be directly or immediately aware of itself in the present moment …" (Strawson. 2009,

177). He claims that " this self-awareness is or can be a constitutive part of experience in such a way that it's rightly construed as a *foreground aspect* of experience even though there's also a respect in which it normally passes unnoticed, being utterly non-thetic" (Strawson, 2009, 177). He describes this direct self-awareness as follows – " It's simply a matter of coming to awareness of oneself as a mental presence – or perhaps simply as *mental presence* – in a certain sort of alert but essentially unpointed, global way" (Strawson, 2009, 179). He suggests that the attainment of such self awareness is an early and rather routine step in certain meditative practices; or it may result from "nothing more unusual than exhaustion, solitude, shock, sex, abstract thought, boredom, or a hot bath" (ibid. 182). He points out that although this experience involves a sense of [narrative] selflessness and is often characterized impersonally as awareness of awareness, or consciousness of consciousness, it is a genuine case of "awareness on the part of the *subject of experience,* of *itself* as it is in the present moment of experience" (ibid. 180).

Similarly in instances of severe pain we are almost exclusively focused on ourselves as the subject of present-moment experience. According to the Lane Medical Library Comparative Pain Scale, severe pain "completely dominates your senses" so that you "can no longer think [about anything else] clearly at all". This shows how severe pain can screen out everything except ourself as the subject who is experiencing the pain.

Sometimes we have an experience of direct self awareness as we wake up. Normally when we wake up in our own bed at our usual time we start our day as routine. Sometimes however, if we are in an unfamiliar situation, sleeping at an unfamiliar time or just waking from a particularly deep sleep, we may experiece a kind of disorientation – we may not know where we are or even who we are – we just know that we are. In these first moments of waking all we have is the feeling of being alive - as Marcel Proust (2003, 9) describes it – "the sense of existence as it may quiver in the depths of an animal". Normally this situation only lasts a few moments and soon our knowledge of our location and our memories return.

Evan Thompson (2015, 236) argues that this scenario reveals two kinds of self-experience –

1. the bodily self-experience of being alive in the present moment, of being sentient, and
2. the autobiographical self-experience of being a person with a story line, a thinking being who travels mentally in time.

In this special waking situation the two kinds of self-experience that are normally intertwined, come apart. Thompson describes the first stage as "an intimate and immediate self-awareness" of "the subject of present-moment experience". This is the element of self that we know by being it.

The above cases may be special instances where this element is isolated, but I claim that this self-experience of being alive is an everyday experience. While not always attended to, it is a ubiquitous element of our stream of consciousness. Sometimes when acting from habit or internalized skill we may temporarily not be consciously aware of ourselves as the experiencer or actor. In many situations however experiences are like something *for me* and actions are *my* actions. And this sense of *me* and *my* is entirely individual; it is not just a matter of first-person givenness. As Wolfgang Fasching (2009, 138) puts it "my experience is not *some* mineness but *my* mineness." I will return to this idea in the next chapter in section 2.4.

Strawson argues that the model of awareness used by the proponents of the elusiveness thesis (such as Ryle) is "too rigid, in so far as it pushes the subject – the 'now-subject', one might say – into being necessarily distinct from its (attempted) object – itself" (Strawson, 2009, 179). The images of an eye that can't see itself or a fingertip that can't touch itself, do express a truth about the limitations of a certain particular sort of thetic, object-posing self-apprehension. But Strawson argues that this is fully compatible with the claim that there's another non-thetic form of occurrent self-apprehension in which the subject can be directly aware of itself in the present moment. This makes our knowing-by-being of ourselves direct but non-explicit. As Bernard Lonergan put it –

> "Objects are present by being attended to, but subjects are present (to themselves) as subjects, not by being attended to, but by attending. As

> the parade of objects marches by, spectators do not have to slip into the parade to be present to themselves" (Lonergan, 1967, 226).

My contention is then that prior to any reflection, self-awareness is the mode of being that we find ourselves living in. This experience of ourselves as agents and subjects, acting in and explaining the world is something that we know by living though it. The experience of knowing by being that which is known, is a condition of the possibility of all other knowing.

In the next chapter I argue that this experience of ourselves indicates the existence of a metaphysical self.

CHAPTER 2. PHENOMENOLOGY & THE METAPHYSICS OF SELF

2.1 Pre-reflective self-consciousness
2.2 The Phenomenology of self
2.3 From Phenomenology to Metaphysics
2.4 My Mineness
2.5 A Bodily Self
2.6 Knowing-by-being a bodily self

The concept 'knowing-by-being' aims to capture the fact that we become aware of ourselves existing/acting/explaining in a world that is in some way separate from us. We have a tacit but direct awareness of ourselves as selves just by being.

This idea has similarities to views in the phenomenological tradition. Shaun Gallagher (2012, 1) states that phenomenology "begins with the fact that we, as agents who must act, and as thinkers who try to get a grasp on what we are doing, are always *already situated* in the world". And Dan Zahavi (2008, 11) points out, that all the major figures in phenomenology defend the view that the experiential dimension of consciousness is characterized by a tacit self-consciousness. This on-going tacit self-consciousness built into experience is thought of as primary or basic, in the sense that it necessary for any explicit reflective self-consciousness.

In this chapter, I examine whether this phenomenological pre-reflective self-consciousness indicates the existence of a metaphysical self. Some contemporary phenomenologists, such Zahavi and Fasching, minimize the concept of self so that it is no more than the way that experience occurs (Zahavi, 2008), or the dimension in which it occurs (Fasching, 2009). I argue however that such a minimal self, confined to the level of experience, is not adequate to capture our awareness of our selves from the way we find ourselves living. Rather, we need to look to what gives rise to experience. I agree with Henry and Thompson's (2011) view that we are prereflectively aware of ourselves as a living body. Knowing-by-being a bodily subject

justifies a more substantial concept of self than phenomenologists such as Zahavi allow.

2.1 PRE-REFLECTIVE SELF-CONSCIOUSNESS

Pre (or non)-reflective self-consciousness (or awareness) is the idea that we have an implicit first-order awareness of ourselves as the subject of whatever experience we happen to be having. This awareness is claimed to be first-order because it occurs along with, or as part of, the experience. In contrast, we might come to have a second-order awareness of ourselves by means of a reflective, introspective examination of our experience. If you ask me are my feet cold, or to tell you what I was just thinking, I would reflect from a perspective one order up from the experience to report to you. Pre-reflective self-awareness, on the other hand, is an awareness I have from the way the experience occurs before I do any reflecting on this experience. Thus, Shaun Gallagher (2012, 127) characterizes it as "an implicit first-order awareness rather than an explicit higher-order form of self-consciousness".

One claim of all the major figures in phenomenology is that this implicit first-order self-awareness is a ubiquitous feature of all experience. Thus, Husserl argued that self-consciousness does not only occur whenever we pay attention to our conscious life. Rather, consciousness, as such, always already involves a self-appearance or self-manifestation (Husserl, 1923/4, 189, 412). He argued that self-consciousness is a feature characterizing subjectivity as such, no matter what the intentional object of our consciousness might be. Thus he argued, "To be a subject is to be in the mode of being aware of oneself" (Husserl, 1973(a), 151).

Similarly, Heidegger argued that the self is disclosed simultaneously with every instance of intentional directedness towards worldly entities. Every worldly experiencing involves a certain component of self-acquaintance or self-familiarity. As he puts it - "Every consciousness is also self-consciousness" (Heidegger GA 27: 135).

According to Sartre, consciousness is essentially characterized by intentionality and each intentional experience is characterized by self-consciousness. For Sartre, self-consciousness is a necessary condition for being conscious of something. As he argues in *Being and Nothingness* – " … for if my consciousness were not consciousness of being consciousness of that table, it would then be consciousness of the table, without consciousness of being so. In other words, it would be a consciousness ignorant of itself, an unconscious – which is absurd" (Sartre, 1943/1969, xxviii). He is arguing that part of the very concept of consciousness is that it must be aware of itself. Thus Sartre argues that the mode of being of intentional consciousness is to be *for-itself (pour-soi),* that is, self-conscious. As he puts it – " This self-consciousness we ought to consider not as a new consciousness, but as *the only mode of existence which is possible for a consciousness of something"* (Sartre, 1943/1969, xxx).

The phenomenologist's claim then is that self-consciousness is not only something that arises when we reflect on our experiences, recognize ourself in a mirror, use the first-person pronoun, or identify with our life story; there is also a more basic form of self-consciousness that is evident in the first-person givenness or manifestation of experiential life. Dan Zahavi calls this "the quality of *mineness*, the fact that experiences are characterized by first-personal givenness. That is, the experience is given (at least tacitly) as *my* experience, as an experience *I* am undergoing or living through" (Zahavi, 2008, 16). Phenomenologists, further claim that any explicit higher-order form of self-consciousness is possible only because there is this on-going pre-reflective self-awareness built into experience. Pre-reflective self-consciousness then is primary and counts as the basic primary fact in the proto-ontology that I introduced in the previous chapter.

Some analytic philosophers also support the idea that all consciousness involves self-consciousness. For example, Owen Flanagan (1992, 194) talks about a low-level self-consciousness involved in experiencing my experiences as mine, in contrast to a stronger notion of self-consciousness that is involved in our thinking about our narrative self. And Uriah Kriegel (2004, 200) writes – "It is impossible to think or

experience something consciously without thinking or experiencing it self-consciously, i.e. without being peripherally aware of thinking or experiencing it." He argues that the phenomenal character of any experience (for example looking at the blue sky) involves both qualitative character (the bluish component), and subjective character (the for-me component). He describes the subjective character as that which remains invariant across all phenomenal experiences. Thus he argues that a phenomenally conscious state's qualitative character is what makes it the phenomenally conscious state it is, while its subjective character is what makes it a phenomenally conscious state at all (Kriegal 2009, 2, 58).

2.2 THE PHENOMENOLOGY OF SELF

Although all the major figures in phenomenology defend the view that the experiential dimension of consciousness is characterized by a tacit self-consciousness, they do not agree on what this tacit self-consciousness amounts to, so that there is disagreement about the status and existence of a self.

Husserl (1931/1960), at least in *Cartesian Meditations*, expresses a view that accords with what I express by knowing-by-being. He argues that the self is experienced in direct intuition. Thus he writes "I exist for myself and am constantly given to myself, by experiential evidence, as 'I myself'" (Section 33).

Some more recent phenomenological writers however deny that having a self-experience gives us any knowledge of a self-object. As Shaun Gallagher and Dan Zahavi argue –

> "To have a self-experience does not entail the apprehension of a special self-object; it does not entail the existence of a special experience of a self alongside other experiences but different from them. To be aware of one*self* is not to capture a pure self that exists separately from the stream of experience, rather it is to be conscious of one's experience in

its implicit first-person mode of givenness" (Gallagher and Zahavi, 2010, 7).

Zahavi develops this idea by claiming that the self is not the *owner* of experiences; rather the self is identified with the ubiquitous dimension of first-personal self-givenness that occurs in our changing experiences. This means that the self is defined as the very subjectivity of experience, and is not taken to be something that exists independently of, or in separation from the experiential flow. Zahavi specifically states that self-awareness is not to be understood as an awareness of "an isolated, worldless self", a "self located and hidden in the head", or "a Cartesian-style mental residuum, that is some kind of enclosed and self-sufficient interiority" (Zahavi, 2008, 125-126). Thus, Zahavi (2008, 126) proposes to separate mineness and selfhood so that he says we should replace the traditional phrase "subject of experience" with the phrase "subjectivity of experience". He claims that using this expression –'subjectivity of experience' – makes it clear subjectivity is not something that can be detached or isolated from experience and that it is not something that experience can lack.

As Zahavi explains it in *Self & Other* –

" for-me-ness, or mineness, doesn't refer to a specific experiential content, to a specific *what*; nor does it refer to the diachronic or synchronic sum of such content, or to some other relation that might obtain between the contents in question. Rather, the mineness refers to the distinct manner, or *how*, of experiencing" (Zahavi, 2014, 22).

Zahavi agrees with Mark Rowlands' formulation of mineness as an adverbial modification of the act rather than a property of the object of that act (Rowlands, 2013).

I argue however that mineness is neither of these; rather it is a property of the subject of the act. So while I agree that the 'what' question about mineness doesn't refer to a specific experiential content, this should not exclude the possibility of awareness of ourselves as the subject of experience. In line with this, Martine Nida-Rümelin (2015, 9) advocates a

shift in the framework of the discussion of these issues from one focused on conscious experiences and their properties to a framework centered on the conscious subject and its properties.

There are two possible consequences of Zahavi's concept of mineness –

1. If mineness only refers to how experiences occur and there is no further self beyond experience, then experience is systematically misleading – every experience occurs in a way that suggests that it is an experience of a 'me', but there is no 'me'. This interpretation would seem to put Zahavi in conflict with the central claim of phenomenology that a careful description of the way experience occurs will give us a true picture of reality.

2. Zahavi's focus on the 'how' rather than the 'what' might reflect a more general phenomenological neutrality about metaphysical questions. In the next section I examine the broader relationship between metaphysics and phenomenology to see what implications this might have for the phenomenology of self.

2.3 FROM PHENOMENOLOGY TO METAPHYSICS

Husserl's early writing supports the idea that phenomenology is, and should be, metaphysically neutral. He rejected questions of whether consciousness can attain knowledge of a mind independent reality, and whether there is an external reality as *metaphysical questions,* which have no place in phenomenological epistemology (Husserl, 1984, 26). He describes phenomenology as a *neutral* investigation (Husserl, 1984, 6), and states that the aim of phenomenology is to *describe* and *understand* the structures of knowledge, not to *explain* how knowledge comes about (Husserl, 1984, 27-28). On this view then, the task of phenomenology is descriptive - to describe accurately what is given without getting involved in any metaphysical explanation.

This rather narrow or limited, object-oriented conception of phenomenology however, was succeeded in Husserl's later writing by a much more ambitious reflective transcendental phenomenology. On at least some interpretations (notably Zahavi, 2003, 2017) transcendental philosophy does have metaphysical implications. While Husserl's transcendental phenomenology is not engaged in the construction of purely formal metaphysical hypotheses, it is claimed to be the right way to approach core metaphysical problems: the form and method of metaphysical problems are claimed to have a phenomenological base.

The key innovations in Husserl's transcendental methodology are the epoché and the reduction. These are the philosophical tools of transcendental analysis. Transcendental analysis, as introduced by Immanuel Kant (1781), is an attempt to discover the conditions of possibility of knowledge, which he took to reflect the way the mind must be structured, and thus, to be prior to experience. Husserl claimed that we could actually put ourselves into a transcendental stance or attitude that would allow us to see how our experience of the world is structured.

The epoché is a method of effecting the change in attitude necessary for transcendental analysis – a change from what Husserl calls the natural attitude (our everyday beliefs, judgments, opinions and theories about how things work) to the phenomenological attitude. *Epoché* is a Greek word for suspension of belief and is often expressed as putting things in parentheses or brackets. Husserl's epoché then involves a suspension by bracketing of these beliefs, judgment, opinions and theories. He claims however, that the epoché is not brought about in order to deny, doubt, neglect, abandon, or exclude reality, but only in order to suspend a certain dogmatic attitude towards reality, to allow us to focus on reality as it is given. The epoché therefore involves a change of attitude towards reality and not an exclusion from reality. As Husserl puts it – "a universal transcendental inquiry includes the world itself, with all its true being" (Husserl, 1959, 432).

The second step in Husserl's phenomenological method is the reduction, which involves turning our attention towards the phenomena as they

appear to us. This involves an investigation of reality in terms of its significance and manifestation for subjectivity. This investigation differs from a direct investigation of the real world, but it remains an investigation of reality.

For Husserl, then, performing the epoché and the reduction makes new discoveries possible and therefore should be seen as an *expansion* of our field of research (Husserl, 1970, 154,I, 66). He compares the achievement of the epoché as a transition from a two-dimensional to a three-dimensional life, in that the previously hidden transcendental subjectivity is disclosed as the subjective condition of possibility for any manifestation (Husserl, 1970, 120).

Husserl himself understood his transcendental philosophy as having significant metaphysical implications. He makes clear that "phenomenology indeed *excludes every naïve metaphysics* that operates with absurd things in themselves, but *does not exclude metaphysics as such*" (Husserl, 1931, 38-39). As I will argue later, a metaphysics based exclusively on the contents of objective scientific enquiry is *naïve*, or incomplete if it excludes consideration of the subjective origins of all enquiry including science. Moreover, Husserl understood his transcendental philosophy as the proper way to the core problems of metaphysics – "… all genuine philosophical problems, all metaphysical problems return to the phenomenological base, where they find their genuine form and method, fashioned from intuition" (Husserl, 1962, 253). As Dan Zahavi puts it, Husserl was not "a quietist when it comes to metaphysical issues … Husserl's phenomenology is as much about the world as it is about consciousness" (Zahavi, 2017, Introduction).

Zahavi (2003, 15) argues that one of the most significant features of Husserl's transcendental phenomenology is a rejection of any two-world theory, which separates the world of appearance (the world as it is for us), from the real world (the world as it is in itself) – the world that objective science investigates. As Zahavi explains it –

> "Just as phenomenology is not *merely* concerned with meaning, it is not *merely* concerned with appearance, or to put it differently, meaning is

not *mere* meaning, nor is appearance *mere* appearance. On the contrary, how things appear, what significance they have, is an integral part of what they really are" (Zahavi, 2003, 14).

On this view, then, the reality of an object is not hidden behind the phenomenon; but rather, it unfolds itself in the phenomenon. As Heidegger (1979, 118) put it, it is phenomenologically absurd to say of the phenomenon that it stands in the way of something more fundamental, which it merely represents. Rather, on this view, the world that appears to us, in perception, in daily concerns, or in scientific analysis, is the only real world. Nevertheless, we can still distinguish between appearance and reality but the distinction does not have the metaphysical significance that is often claimed in analytic philosophy –

"Although the distinction between appearance and reality must be maintained, it is according to transcendental phenomenology not a distinction between two separate realms (falling in the province of phenomenology and metaphysics, respectively), but a distinction internal to the realm of appearances. It is a distinction between how the objects might appear at a casual look, and how they might appear in the best of circumstances" (Zahavi, 2003, 14).

Zahavi's own concept of the core or basic self as 'how' experience occurs may be problematical on this analysis. On Zahavi's account 'how' all experience occurs is 'for-me' or in the mode of 'mineness'. If how things appear were an integral part of how they really are, then we would expect that if experience appears as if it is the experience of a 'me', then there is a 'me'. If experience is not systematically misleading on the object side then it is reasonable to expect that it will not be systematically misleading on the subject side as well. Husserl himself argued for the factual existence of transcendental subjectivity – "the eidos transcendental I is unthinkable without the transcendental I as factual" (Husserl, 1973(b), 385).

In *Ideen II,* Husserl argues that reflection on a single experience will reveal the nature of the pure ego, and that it is absurd to claim that this ego doesn't really exist or that it is ultimately quite different from what it is

revealed as (Husserl, 1952: 104). Galen Strawson develops a similar line of reasoning but expresses it in terms of necessary and sufficient conditions –

> "a *necessary* condition on qualifying for the title 'self': nothing can be allowed to count as a self unless it possesses all the properties attributed to the self by all genuine forms of self-experience – whatever other properties it may or may not possess. …. a sufficient condition on qualifying for the title 'self: nothing can fail to count as a self if it possesses all the properties that feature in all genuine forms of self-experience" (Strawson, 2009, 56).

Zahavi endorses this kind of reasoning in that it takes self-experience to be the key to understanding what it means to be a self, but he goes "one step further" –

> " … ultimately [phenomenologists] would go one step further by claiming that when it comes to this most basic form of self, it is constituted in and through self-experience, that is, that the phenomenology of that kind of self just is the metaphysics of that kind of self" (Zahavi, 2014, 17-18).

Zahavi makes clear in a footnote (2014, 18, Footnote 5) however that his reference here to metaphysics refers to the question regarding the existence and reality of selfhood and not to what kind of 'stuff' a self might ultimately be made of. Significantly he adds "… and I am not suggesting that a phenomenological investigation of the self can per se resolve the later question". This suggests that Zahavi understands that there is a metaphysical limit even to transcendental phenomenology. While how things appear reveals what they really are, when it comes to consciousness itself, phenomenology cannot go beyond or behind experience. Zahavi, himself states that some genuine metaphysical questions might be beyond the reach of phenomenology (Zahavi, 2003, 16), and this might be one of them.

On Zahavi's view then the subject side of experience is different from the object side. So while the objects of experience reveal reality, there is

nothing more on the subject side than the way it appears. 'Mineness' is just the way that experience occurs (from a point of view), but it doesn't reveal any 'me' or 'experiencer'. Nevertheless, Zahavi wants to keep the concept of a 'self', but accepts that his concept of a core self is no more than a "formal kind of experiential individuation". He claims that "the very subjectivity of experience, its first-person character, although being quite formal, amounts to a kind of self" (Zahavi, 2014, 23).

However, Jay Garfield questions whether "… a kind of pure subjectivity … a sort of mineness, with no substantial owner, either synchronically or diachronically .. would .. constitute an analysis of selfhood at all, or an abandonment of that very idea?" (Garfield, 2015, 99). As Soren Klausen (2019) argues there is a distinction to be made between different two notions of perspectivity – first, *content perspectivity,* and second, the notion of *having,* or embodying or enacting a perspective. Content perspectivity is the fact that something appears in a certain way. This applies to drawing, photos and films, as well as conscious experience. It resembles what John Searle terms *aspectual shape*: when something is presented to a subject (in thought, perception, desire and so on) it is always presented under a certain aspect, or in a certain way (Searle, 1992, 156f.) The notion of having, or embodying or enacting a perspective, however is not a mode of presentation or appearance; rather it is the very directedness or openness towards something. On this analysis then the subjectivity or awareness component of phenomenal consciousness (as opposed to the qualitative character 'what it is like' component) splits into a phenomenal-content part and an ownership or perspectival part, and as Klausen puts it –

> "[phenomenal consciousness] states may indeed always comprise a feeling of mineness. But they are also *had* by me (or you, or someone else); and the latter property cannot, it seems, be cashed out in terms of the former" (Klausen, 2019, 18).

I agree with this analysis in that an adequate understanding of the subject of experience must include reference to an owner of the experience, and not just to how the experience occurs. As Martine Nida-Rümelin (forthcoming) argues the claim that experiences have subjective character is best

understood "as expressing the metaphysical fact that there is someone undergoing the experience *for whom* it is like something to undergo the experience". She rejects the idea that the metaphysical issue about what it is for an experience to belong to a subject and phenomenological issues concerning pre-reflective self-awareness are separate theoretical issues, which one may treat independently. As she puts it –

> "One cannot expect to be able to develop a satisfying account of pre-reflective self-awareness (of what it is to be *aware* of oneself as the subject of experience one undergoes) without reflecting upon what it *is* to be the subject of that experience since the latter is precisely what one is aware of in the former."

Zahavi (2014, 18) himself sees his concept of a core self as occupying the middle position between two opposing positions. On one view, "the self is some kind of unchanging soul-substance that is distinct from and ontologically independent of the worldly objects and conscious episodes it is directed at and of which it is the subject". According to the other view "there is nothing to consciousness apart from a manifold or bundle of changing experiences". Zahavi claims that the advantage of his core or experiential self, is that it is not a separately existing entity, but neither is it simply reducible to a specific experience or subset of experiences. But as Søren Klausen (2019) points out, there is an "essential tension" in Zahavi's view in that it vacillates between conceiving of mineness as a phenomenal feature and denying it to be so. Klausen says that this tension is particularly exemplified in the way that Zahavi (2014, 22-24) explains the thought experiment of Mick and Mack – two perfect twins who are imagined to be both staring at a blank wall. As Klausen observes –

> "Zahavi describes Mick and Mack as phenomenologically type-identical, yet at the same time insists that each of their experiences will have its own distinctive manifestation, and so that Mick's experience will be quite unlike Mack's" (Klausen, 2019, 16).

Zahavi characterizes the difference between Mick and Mack's experience in terms of them each having their own "distinct pre-reflective self-manifestation" (Zahavi, 2014, 24). It is difficult to understand how Zahavi

can characterize the 'manifestation' of the self as constituting the core notion of self when experience itself only occurs to individuals from *their* first person perspective.

Moreover, Zahavi's characterization of the substantial self is rather extreme and not the only possible understanding of a substantial self. Even an intuitive understanding of our selves is not necessarily as "unchanging", "soul-substance", or "ontologically independent of … worldly objects". It seems also possible to understand our selves precisely as constantly changing, material substances that are ontologically dependent on other worldly objects. In the last section I examine one such proposal. But first, I will examine Wolfgang Fasching's criticism of Zahavi's core self and his useful exploration of mineness.

2.4 MY MINENESS

Zahavi presents his concept of a core or experiential self as "formal kind of experiential individuation" – experiences occur from individual perspectives. On this view, it seems that experiences take place just by being first-personally given without thereby presupposing a 'first person' - experiences are their own 'stage' of appearance. Wolfgang Fasching (2009) points out however that mineness is not just an anonymous individuation of experience: rather, each 'mineness' is unique. As Fasching puts it -

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

Fasching presents his concept of 'my mineness' in contrast to Derek Parfit's reductionist view that "my experience" means nothing more than "*this* experience, occurring in *this* particular life" (Parfit, 1987, 517). On this view 'my' is simply an indexical mode of reference, and the meaning of all indexical expressions, including 'I', can be explained through the meaning of 'this', namely in its 'self-referring use'. However Geoffrey Madell argues that you cannot explain away the peculiarity of the first-person pronoun as being merely a case of indexical reference. Rather it is the other way around - all indexical expressions parasitically refer to 'I' (Madell, 1981, 39-40). Thus, 'here' is where I am, 'this' is what I am pointing at etc. As I mentioned in the introduction, John Perry (1979) calls 'I' the essential indexical because if we replace it with other designations we no longer have an explanation of my behaviour. So, although there are many indexicals, there is only one *essential* indexical.

Fasching agrees with this and argues for the uniqueness of 'I' by means of two counterfactual thought experiments that involve the possibility of 'transferability' and 'separability' –

1. "Sometimes I might wish to be someone else (i.e. another person, with totally different experiences and even different character traits)". Fasching argues that although this wish is clearly conceivable and not contradictory, it just in fact is not the case. But this could not be possible if 'I' were to refer indexically to the stream of consciousness the present (wishing) experience belongs to. How could this stream of consciousness be another stream of consciousness? As Fasching puts it – "that I can imagine being someone else seems to be incompatible with the logic of 'this'". Fasching's transferability intuition suggests that I am not totally made up of all the experiences I have and the character traits by which other people know me. Rather there seems to be an ineliminable 'me', which could have totally different experiences and different character traits, but still be 'me'.

2. I can "wish (or imagine) that the very person I am (with all his physical and mental properties) would not be *me.*" Fasching argues that we have an intuition that there is a logical contingency in the fact that I experience the

experiences I actually do - "The person I am with all his experiences could have existed, doing what I do, experienc[ing] what I experience, with the only difference being that he would not be me. Nothing would have been different for anyone, except for me: I would not have existed." As in the first case, there is no analogy in the case of 'this'. It makes no sense to say 'This sentence could exist with all its properties except that it would not be this sentence'. Thus, Fasching is arguing for the possibility of the separability of 'me' from all my physical and mental properties.

Fasching argues that the basis of the difference between 'this' and 'I' is the intuition that there is a logical contingency in the fact that I experience the experiences I actually do i.e. in the fact that '*this* stream of consciousness' is '*my* stream of consciousness'. This contingency intuition gives rise to what Herbert Spiegelberg (1961) termed the 'I-am-me' experience, which mainly occurs in childhood or adolescence. This experience involves an unsettling puzzlement about one's own being oneself – about the fact that it is oneself who one is. It appears to be a *contingent* fact that one is the person one is. As Thomas Nagel argues, it appears that 'I am TN' cannot be an *identity* statement, but rather is a subject-predicate proposition (Nagel, 1986, 55). Tsuneo Watanabe argues that 'I-am-me' experiences involve an experiential disruption of the self-evidence of the self. When this occurs there is a coming apart of the subjective world with I as the center, and the objective world in which I am just one of many selves (Watanabe, 2009, 12,19).

Fasching argues that in the case of 'I', the subjectivity of experience involves a further dimension that allows for a difference over and above all objective facts, namely the subject itself. As he puts it –

> "That I am me (the person that I am) is a 'further truth' over and above *all* objective truths (be they physical or mental)" (Fasching, 2009, 141).

Fasching's concept of 'further truth' expresses the realization that there is something more we know that is more basic or primary than any objective truth; this is in line with what I, (from Shaun Gallagher), term 'primary facts'. Fasching takes the idea of a 'further truth' from Thomas Nagel who argues that 'being me' is a truth that can be stated and

understood only from my perspective, in the first person, because there is no room for such a truth in an objective, centerless world (Nagel, 1986, 56). Nagel concludes that indexicals are ineliminable from a complete conception of the world (ibid. 57); as Perry (1979) pointed out, 'I' is the essential indexical.

Fasching thus argues that the 'I' or 'self'' is *presupposed* by the very existence of conscious experiences (as that which 'has' them), rather than that it is, as Zahavi argues, in some way *constituted* by them. This would seem to support the idea of knowing-by-being which suggests that we are aware of ourselves not just as an ever-present way in which everything is experienced, but also more directly as the source of, or recipient of, any possible experience. We are aware of ourselves as the locus of experiences – we are what has experiences – we are the experiencer, prior to any particular experience.

This would make the self an important feature, arguably the most important feature, of any possible ontological description of the world. Whatever else exists, we are sure at least that selves exist because we are one. We know ourselves as the kind of entity that can act in and produce theories about the world.

However, although Fasching argues for the individualness of mineness, he does not advocate the self as an entity separate from experience. Rather he says that –

> " I *am* my consciousness, and not some entity that has or owns it …'the I' has to be conceived of as having the character of a *dimension* with regard to its experiences … [a] dimension of first-personal presentation which has no other essential property than being me: a dimension that is completely 'subjective'" (Fasching, 2009, 144, 146).

It is not clear however, how he reconciles individual mineness with a universal feature such as a dimension, and the unwillingness to look beyond experience seems an unnecessary restriction. In the next section I examine a proposal that would reunite subjectivity and selfhood in a bodily self.

2.5 A BODILY SELF

Not all phenomenologists have such a minimalist notion of self as Zahavi or Fasching. For example, Henry and Thompson (2011) propose an account of consciousness that challenges attempts to separate subject and self. Interestingly, they present their view in contrast to Mira Albahari's Buddhist inspired overtly no-self view of subjectivity (Albahari, 2006), but I think their argument can equally to apply to Zahavi's core self view.

Albahari uses interpretations of early Indian Buddhism to argue that the self we usually take ourselves to be is an illusion. I will consider the Buddhist no-self view and illusionism in more detail in chapter 6. As Albahari explains it - while our everyday sense of self is as a bounded subject that is a personal owner of experiences and a controlling agent of actions, in reality there is no such self. The impression of a self arises from a process of identification with transitory mental and bodily experiences as 'me' or 'mine'. However, Albahari does affirm the experiential reality of a 'witnessing subject'. Thus, she makes a distinction between a mere 'subject' of experience and a self. Being a subject of experience is simply a matter of instantiating an impersonal perspectival awareness of the world. The experience of selfhood additionally involves 'boundedness', and a sense of unique 'personal ownership' or 'my-ness'.

In effect, Albarhari's 'witnessing subject' and Zahavi's experiential or core self are very similar. Albarhari asserts the reality of the 'witnessing subject' but denies that it has any attributes of a self; Zahavi asserts the reality of a core self that is entirely constituted by the subjectivity of experience.

Hubert Wykretowicz (2018) claims that Gallagher and Zahavi's (2008) minimal self is a sort of limit case of the self rather than the self itself. He point's out that Edmund Husserl (1960, 35) thought of the self as a unified phenomenon, a 'gestalt' that is hard to reduce to neuronal patterns. He talks of the ego or self as a *habitus* - a set of dispositions, habits, tendencies, skills

that remain with me however unaware of them I am (Husserl, 1996). Wykretowicz uses these ideas to argue that "the core of my identity is not a minimal self". He argues that the self is not only a temporal formal structure of the stream of consciousness. Rather, he argues "my self is always already a structure that is 'filled up' with the dispositions, skills and tendencies that develop out of my experience of the world, and which constitute for me a set of beliefs, preferences and capabilities that are always already shaping my own experience before I know it." This is what Husserl (1996a: 58, 60d) calls "a person", a "type" or a "style of living". On this view then what it feels like to be someone "must be understood, first and foremost, as the specific quale of my being a habitus, that is living within a set of beliefs, preferences and skills that are deeply rooted in my body". For Wykretowicz then self-identity is found not simply in brain correlates but in "my body as a whole, which has been shaped by biology, society and self-experience."

As Henry and Thompson argue - in order to be a subject of experience even in the minimal sense of 'witnessing-from-a-perspective', one must be prereflectively aware of oneself as a living body, i.e., one must be a bodily subject. They distinguish their view from the view that the subject of experience is embodied, which implies that the body is an object that belongs, in some special way, to an essentially mental subject. Rather, their view is that the subject is constituted bodily (Henry and Thompson, 2011, 230).

Descartes went some way towards this position by arguing that one's soul (or self) is not located in one's body as a pilot within a ship but instead is "very closely joined" and "intermingled" with it, so that the two "form a unit" (Descartes 1986, 56). Nevertheless, for Descartes soul and body remain two entities, and it is precisely the undeniable fact that they do form a unit, that Descartes finds hard to explain. Merleau-Ponty, on the other hand, rejects this dualism, claiming that one's self is not merely *embodied,* but rather *bodily* –

> "But I am not in front of my body, I am in my body, or rather I am my body" (Merleau-Ponty, 1962, 150).

Merleau-Ponty does not understand "I am my body" in a materialist way, as meaning I am (or my self is) nothing more than a complex physical object: rather he argues that I am a *bodily subject*.

As Husserl (1973) puts it the body for human beings is experienced in a double way. It is both a physical, organic object in the world that could be perceived as such from the outside (*Korper*), but it is also lived and experienced from the inside, within an irreducible, first-person perspective (*Lieb*). And, as I mentioned in chapter 1, for Husserl, subjective experience has an intrinsic temporal dimension – events flow from the past through the present towards the future. To be a subject, he says, is to 'live through' something (*Erleben*), rather than just to survive or stay alive, as a merely organic creature (*Leben*). Crucially for Husserl the body as an object depends on the lived body –

> "… in all experience of things, the lived body is co-experienced as a functioning lived body (thus not as a mere thing), and … when it itself is experienced as a thing, it is experienced in a double way – i.e. precisely as an experienced thing and as a functioning lived body together in one" (Husserl, 1973, 57).

As Dan Zahavi explains it, the constitution of the body as an object is not an activity that is carried out by a disincarnate subject. Rather it is a self-objectification brought about by an embodied subject. Thus he argues that – "The lived body precedes the perceived body. Originally, I do not have any consciousness *of* my body. I am not perceiving it. I *am* it" (Zahavi, 2019, 83).

Merleau-Ponty explains his conception of bodily subjectivity by means of a contrast between the perceptual unity of a physical object through its perspectival variations and the experiential unity of one's own body. A physical object, such as cube, stands before me and presents various facets to me. What unites these facets as facets of one and the same cube, is the law of their perspectival variation. My body, however, does not stand before me – so what unites the various experiences I have of my body is not a law of perspectival variation, but rather an integrated perceptual and motor "background of significance", "style", or "bodily bearing" –

> "What unites 'tactile sensations' in the hand and links them to visual perceptions of the same hand, and to perceptions of other bodily areas, is a certain style informing my manual gestures and implying in turn a certain style of finger movement, and contributing, in the last resort, to a certain bodily bearing" (Merleau-Ponty, 1962, 150).

Further, he argues that the mechanics of the skeleton cannot, even at the scientific level, account for the distinctive positions and movements of my body.

This understanding of one's bodily subjectivity as constituted by a certain style or bearing leads Merleau-Ponty to liken the body to a work of art –

> "The body is to be compared not to a physical object, but rather to a work of art … A novel, a poem, picture, or musical work are individuals, that is, beings in which the expression is indistinguishable from the thing expressed, their meaning, accessible only through direct contact, being radiated with no change of their temporal and spatial situation. It is in this sense that our body is comparable to a work of art. It is a nexus of lived meanings, not the law for a certain number of covariant terms" (Merleau-Ponty, 1962, 150-151).

Thus Merleau-Ponty argues that "I am my body" in a way not unlike the way that an artwork is what it expresses. My body is an expressive being, and what it expresses is my subjectivity. The Norwegian novelist Karl Ove Knausgård (2018) illustrates this through the example of how we can recognize someone from the way they walk. He says that describing the motor function – putting one foot in front of the other – says nothing about what it feels like to walk and even less about the feeling of watching others walk. In the case of a person you've known for a very long time -

> "… their entire character seems to be contained in the way they walk when they come towards you. You see it at once, here 'he' or 'she' comes. In that unique pattern of movements is all that you know and have experienced about that person, though not as separate, clearly divided

parts, what you see is in a way the sum of that person, what they 'are' to you. They come walking, you see them, and that's it" (Knausgård, 2018, 173-4).

Shaun Gallagher (1986, 1995) analyzes our bodily being in terms of a distinction between the conscious "body image" and the unconscious "body schema". The body image is the body as an intentional object of consciousness; so, for example it is what you experience by looking directly at it or at one's reflection in a mirror. As Dorothee Legrand (2006) puts it – it is consciousness of the body-as-object. In the body image, the body is experienced as owned by the experiencing subject, and the image is typically a partial representation in that conscious awareness can only attend to one part or area of the body at a time.

The body schema, on the other hand, is an integrated set of dynamic sensorimotor processes that organize perception and action in a subpersonal and unconscious manner. Gallagher makes clear that the body schema is not phenomenologically available to the subject –

> "The body schema … is not the perception of 'my' body; it is not the image, the representation, or even the marginal consciousness of the body. Rather, it is precisely the style that organizes the body as it functions in communion with its environment" (Gallagher, 1986, 549).

Gallagher's 'body schema' does capture some of what Merleau-Ponty had in mind in terms of 'style'. However, Dan Zahavi (1999, 98, 240) argues that the distinction between body schema and body image leaves out a fundamental form of bodily experience, namely, prereflective bodily self-consciousness. So while most of the time one's body is not present as an intentional object, it is nevertheless experienced in an implicit, tacit, and prereflective way. This kind of experience is consciousness of the body-as-subject. It corresponds to the relation of the lived body to itself, that is, to one's experience of one's body as perceiving and acting rather than as perceived. Sartre (1956, liii) describes this sort of self-consciousness as 'non-positional' or 'non-thetic' because it does not posit one's body as an object; Merleau-Ponty calls it prereflective.

Evan Thompson endorses the term 'prereflective' because it has both a logical and temporal sense –

> "Prereflective experience is logically prior to reflection, for reflection presupposes something to reflect upon; and it is temporally prior to reflection, for what one reflects upon is a hitherto unreflected experience" (Thompson, 2010, 250).

Thompson explains how prereflective bodily self-consciousness is evident in touch – "not only do we feel the things in touch, but we feel ourselves touching them and touched by them" (ibid, 250). Usually one's body does not sense itself explicitly because one is focused on the perceptual object, not oneself. But one's body can also sense itself explicitly, as when one hand touches the other. In this situation, the one touching is the thing touched, and the thing touched senses itself as the one being touched. Merleau-Ponty argues that this form of bodily self-consciousness makes vividly apparent the lived body's unique status as a physical subject – "… the body is a 'perceiving thing', a 'subject-object' (Merleau-Ponty, 1964, 166).

Henry and Thompson (2011) argue that the phenomenon of prereflective bodily self-awareness entails a basic experience of boundedness between one's subjective body and the phenomenal world with which one is bodily engaged, so that the processes responsible for one's egocentric bodily perspective on the world are simultaneously responsible for the formation of a minimal bounded self in the form of a living bodily subject of experience (Henry and Thompson, 2011, 231).

This boundedness of the self, however, should not imply that the self is an unconstructed entity that exists independently of its experiences and the rest of the world. But the fact that our self may be constructed does not make it an illusion. As Henry and Thompson argue, the distinction between self and world is best understood as a mutually specifying relation that is enacted in the process of living. Also, the boundary is not fixed or inflexible. Rather the felt boundary between subject and world is plastic and fluidly adaptive (Henry and Thompson, 2011, 248).

As we have seen, some phenomenologists argue that bodily self-experience belongs on the subject side of the subject-object distinction. In order for it to be possible to perceive a world of objects (including the body-as-object), one must be a bodily subject. This approach would lead to a modification of the Kantian idea of transcendental subjectivity, such that the transcendental and empirical perspectives are to be conceived of as intertwined and complementary perspectives. As Robert Sokolowski puts it "The empirical and the transcendental egos are not two entities; they are one and the same being, but considered in two ways. Moreover, it is not just our manner of considering the ego that introduces the distinction … Rather the ego exists in this double manner" (Sokolowski, 2000, 113). This would accord with Husserl's (1931/1960, 130) claim that transcendental subjectivity must necessarily conceive of itself as a worldly being if it is to constitute an objective world, since objectivity can only be constituted by a subject, which is both embodied and socialized.

2.6 KNOWING BY BEING A BODILY SELF

In this chapter I have been critical of Zahavi and Fasching's concept of the self as being too minimal in that it restricts the core self to the way that experience occurs or the dimension in which it occurs. Zahavi accepts that there are also far more complex forms of selves (such as the narrative self) but he claims that the experiential notion remains fundamental in the sense that nothing that lacks this dimension deserves to be called a self. As Gallagher and Zahavi (2008, 204) explain it – "the crucial idea is that some minimal form of self-*experience* is essential for selfhood". They make this claim in contrast to Dan Dennett's (1991, 174, 414) observation that no organism can survive or act without being able to distinguish between self and nonself, even if only on a non-conscious level. Gallagher and Zahavi claim however that according to the phenomenological conception of the

minimal self, selfhood requires more than merely a non-conscious differentiation between oneself and the environment – it requires conscious self-experience.

But some phenomenologists such as Thompson (2007) have a wider ontology of selfhood, with selves occurring at different levels of complexity, and the most basic kind involving the simplest biological organisms separating themselves from their environment in the process of autopoesis. I mentioned this idea in Chapter 1, and will return to it again in Chapter 6. Even at the level of self-experience that Gallagher and Zahavi are talking about, I argue that the experience we prereflectively have of ourselves is not confined to the level of experience alone, but precisely involves an experience of a bodily subject situated in the world.

Transcendental phenomenology (at least as presented by Zahavi) is reluctant to go beyond the experience of subjectivity to consider what gives rise to it. This is an unnecessary restriction. If what appears as an object of consciousness is a guide to reality, then the way that we are aware of consciousness itself could guide us to an understanding of what gives rise to consciousness – 'mineness could reveal a 'me' or 'self'. Also, as I argued in Chapter 1, not all our mental states are intentional, that is, about something. Sometimes, I argue we can be directly aware of ourselves as subjects of any possible experience or none. Transcendental phenomenology's field of study could be expanded to include this.

The challenge for an adequate theory of self is to fully account for our pre-reflective sense of self in such a way that is compatible with the empirical perspective but does not deny or minimize our self as we know it to be, by being it. If we are not mistaken that we can explain the world and act autonomously, then our concept of self must accommodate these capacities. To say that selfhood is more substantial, however is not to imply that it is a process-independent ontological substance that could exist all by itself, in isolation from the rest of the world. But neither does it make sense to say that we exhibit subjectivity without being a self. To reject the existence of self, or to claim that selfhood is no more than subjectivity, while endorsing

the reality of subjectivity is to miss out on what subjectivity really amounts to. Selfhood is significant or substantial in that it is the source of our explanations of, and actions in, the world, and this is the self that we know by being it.

If, as I have argued in this chapter, we are bodily subjects or selves in the world then we might expect that empirical science would have something to say about selves. This is the topic of the next chapter.

CHAPTER 3. PHYSICS, PHYSICALISM & CONSCIOUSNESS

3.1 Worldviews and a Theory of Everything
3.2 The Scientific Picture of the World
3.3 Physicalism
3.4 Physics and Mathematics
3.5 The Interpretation of Quantum Physics
3.6 The barrier of consciousness and the burden of proof.
3.7 Escape routes
3.7.1 Watchful waiting
3.7.2 Embrace emergence
3.7.3 Favour fundamentality
3.8 Towards modifying metaphysics

3.1 WORLDVIEWS AND A THEORY OF EVERYTHING

Wilfred Sellars characterized the aim of philosophy as "… to understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term" (Sellars, 1963, 1).

In Chapter 1, I suggested that Wittgenstein's phrase "The world as I found it" expresses the idea that we find ourselves already having a pre-reflective understanding of such a worldview – what Edmund Husserl called the lifeworld. For Husserl the lifeworld is 'pre-given' – it is always already there for you as a horizon in which things are given, and the meaning of the given things is intersubjectively constituted – 'all of us together' are in the life world.

Wilfred Sellars introduces a somewhat similar idea with his concept of 'the manifest image', which he characterizes as "the framework in terms of which man came to be aware of himself as man-in-the-world" (Sellars, 1963, 6). It is the framework in terms of which we ordinarily observe and explain our world by making a fundamental distinction between persons and things.

For both Husserl and Sellars, science presents a challenge to our human worldview. For Husserl science is an abstraction away from the lifeworld. I will return to phenomenological views of science in more detail in chapter 5. For Sellars there is a "clash" between "the 'manifest' image of man-in-the-world" and "the scientific image." These two 'images' are idealizations of distinct conceptual frameworks in terms of which humans conceive of the world and their place in it. The manifest image is methodologically prior to the scientific image because it provides the initial framework in which science develops. Nevertheless the scientific image claims explanatory superiority. As Sellars puts it –

> "the scientific image presents itself as a *rival* image. From its point of view the manifest image on which it rests is an 'inadequate' but pragmatically useful likeness of a reality which first finds its adequate (in principle) likeness in the scientific image" (Sellars, 1963, 20).

William Seager (2012) calls the worldview of science – 'the scientific picture of the world'. Unification of theories, within that scientific picture, has been a motivating aim of science, and for many scientists the ultimate goal is to produce 'a theory of everything'. As John D. Barrow explains it, the dream of modern scientists is to discover "a monumental 'Theory of Everything' which will unite the laws of Nature into a single statement that reveals the inevitability of everything that was, is and is to come in the physical world" (Barrow, 2007, Preface).

In this chapter I examine whether 'the scientific picture of the world' can be 'a theory of everything', and if it can't, what it leaves out.

3.2 THE SCIENTIFIC PICTURE OF THE WORLD

Science is generally thought to be the most reliable or systematic way of finding out about the world. If we want to know whether our planet is getting warmer, or if genetically modified crops are safe, we rely on scientists to find the answers. Science supplies answers to specific questions.

Some philosophers, and particularly metaphysicians, however, want to find the answers to very broad questions about the overall way everything happens – "how things … hang together" as Sellars put it. This may not be a question that is a concern of all scientists. Many are interested in examining some aspect of the world in great detail. But at least some scientists claim to be interested in finding a 'theory of everything', and assert that objective science is *the* way to find the truth about the world. In the 1990's the Nobel laureate physicist Sheldon Glashow asserts this in the form of a scientific creed –

> "We believe that the world is knowable: that there are simple rules governing the behavior of matter and the evolution of the universe. We affirm that there are eternal, objective, extra-historical, socially-neutral, external and universal truths. The assemblage of these truths is what we call science, and the proof of our assertions lies in the pudding of its success."

Some scientists even claim that we already have a theory of everything at least in broad outline. This is what one-scientist claims we know –

> "The picture of the world that we have finally reached is the following. Some ten or twenty different quantum fields exist. Each fills the whole of space and has its own particular properties. There is nothing else except these fields; the whole of the material universe is built of them. Between various pairs of fields there are various kinds of interaction. Each field manifests itself as a type of elementary particle. The number of particles of a given type is not fixed, for particles are constantly being created or annihilated or transmuted into one another. ... it remains perpetually astonishing that our solid world of trees and stones can be built of quantum fields and nothing else" (Dyson, 1953, 62).

Sean Carroll writing in 2010 seems to have less astonishment of the difference between 'our solid world' and the scientific explanation of 'everything', but is equally sure that physics can account for everything -

> "The laws underlying the physics of everyday life are completely understood ... All we need to account for everything we see in our everyday lives are a handful of particles – electrons, protons, and neutrons – interacting via a few forces – the nuclear forces, gravity, and electromagnetism – subject to the basic rules of quantum mechanics and general relativity" (Carroll, 2010).

William Seager (2012) agrees that science has found "rich interconnections which underpin ever more comprehensive explanatory projects". Science has been astonishingly successful in constructing a scientific picture of the world that is comprehensive, unified and yet, hierarchical. According to this scientific picture, the world began in a relatively simple and purely physical state, constituted by a small number of simple physical entities interacting according to the fundamental laws of physics. The general physicalist project is to give an account of how the features of the world, which we now observe, emerged from the initial conditions of the Big Bang. Thus, Max Tegmark argues that features such as consciousness and conscious persons must ultimately be explained in terms of arrangements of fundamental particles –

> "I approach this hard problem of consciousness from a physics point of view. From my perspective, a conscious person is simply food, rearranged. So why is one arrangement conscious, but not the other? Moreover, physics teaches us that food is simply a large number of quarks and electrons, arranged in a certain way. So which particle arrangements are conscious and which aren't?" (Tegmark, 2017, 284-285).

According to Seager, the hierarchical features we experience are understood in terms of '*conservative*' (weak) emergence; all emergent features are strictly and entirely determined by the nature and state of the fundamental physical features of the world. This picture may not yet be

complete but "[no] permanent dead end has yet blocked our progress and everything seems to hang together remarkably well" (Seager, 2012).

Seager argues that there are three interlocking features that are central to the scientific picture of the world – completeness, closure, and resolution. Completeness is the doctrine that everything in the world is physical and as such abides by closure and resolution. Closure entails that there are no 'outside forces' – everything that happens, happens in accordance with fundamental physical laws so as to comply with resolution. Resolution requires that every process or object be resolvable into elementary constituents, which are, by completeness, physical and whose abidance with laws governing these constituents, leads to closure (Seager, 2012, 156-157).

The success of science in giving a physical explanation of the world has led to the adoption of physicalism as the dominant metaphysics of nature in current times.

3.3 PHYSICALISM

William Seager (2014, 143) states that "Physicalism is the view that everything that exists is ultimately physical". This view is metaphysical in that it is a view about everything. It is a monist view in that it claims that there is only one basic kind of reality; and it is a specific view in that it claims that reality is ultimately physical. The qualification 'ultimately' can be understood as a recognition that the theory may not be immediately obvious – everything as it appears to us is not obviously physical – so the theory will have to give some explanation for this.

But this definition does not tell anything about what it is to be physical. Seager approaches this by saying that in the first instance physical reality is known by ostension – we are in perceptual contact with aspects of the world, which are paradigmatically physical – and we begin there. This might suggest that the physical is "continuously extended, space filling and exclusively space occupying stuff" (Seager, 2014, 144). But although

physical reality is initially known by ostension, it is more fully revealed by science and most intimately by the foundational science of physics.

A difficulty for our initial intuition of physical reality is that science has revealed that the physical is much stranger and much less material than our everyday experience of the world. And, as I argue in Section 3.4, science has not yet arrived at a settled opinion on what is fundamental, so physicalism may not necessarily mean 'materialism'. Also if what entities exist is to be determined by physics alone, Carl Hempel (1969) points to a dilemma –

- if the physics at issue in the physics based account is current physics, then the resulting physicalism will surely be false; for current physics is surely both incomplete and to some extent inaccurate.
- if the physics at issue is rather future (or ideal) physics, then the resulting physicalism will be either indeterminate in content, trivially true, or compatible with entities that are intuitively physically unacceptable.

Jessica Wilson (2005) points out a problem with the last possibility -"If physics were to posit fields or particles themselves possessing mental properties as opposed to composing or constituting complex entities possessing such properties), then physicalism would be or should be thereby falsified".

But Wilson claims that physicalists have not handed over *all* authority to physics to determine, *a posteriori*, what is physical. "Reflecting the historical roots of physicalism in materialism, as foundationally committed to understanding mentality as nothing over and above complex material goings on, one feature has remained definitive of the term "physical" (as this term enters into formulating physicalism, at any rate); namely, that physical entities are not fundamentally mental; physical entities do not individually either possess or bestow mentality". So Wilson proposes that " an entity is physical just in case it is (approximately accurately) treated by current or future (at the end of inquiry, ideal) physics, and is not fundamentally mental".

This seems to make physicalism a peculiar project – it involves giving a blank cheque to one scientific discipline (an essentially mental process) to find the fundamental features of reality – with the stipulation that these

features do not possess or bestow mentality (which is necessary to carry out the project).

In spite of these possible problems, physicalism has become the dominant metaphysical position of the modern era because of its unparalleled scope, scale and explanatory power. Seager points out that vast swathes of the world have fallen under the explanatory project of physicalism and the scientific enterprise so that it seems that we are now close to having a comprehensive physicalist picture of the universe. He says that the physicalist picture is "intellectually compelling in its scope, explanatory power, metaphysical simplicity and ontological elegance" (Seager, 2014, 148). This success has led to what Seager terms "the optimistic induction" (Seager, 2014, 150). This is based on the fact that for over four hundred years now the scientific enterprise has enjoyed an on-going pattern of success based on the underlying presupposition that there is a physicalist account of all phenomena. There were many times when this success appeared to be threatened or stalled by some seemly recalcitrant element, such as vitalism. Ultimately, however, these recalcitrant elements yielded to a physicalist understanding and integration with the larger scientific picture of the world. The optimistic induction is the inference that in the face of such long term and extensive success, scientific metaphysics will be completed.

Seager points out that the physicalist worldview has progressed by the revelation of the mechanisms of ontological dependence via the exhibition of reductive epistemological dependence. The successes of physicalism "have shown us how we could understand some (relatively) 'macro' phenomenon via an understanding of some (relatively) 'micro' phenomenon" (Seager, 2014, 151). So while the pessimistic induction points to the evident historical fact that all our scientific theories except our current ones (so far) have turned out to be false (Laudan, 1981), the optimistic side, as Seager argues, is that this history of 'failure' has never revealed a problem with the physicalist presupposition behind the replacement of one theory with another one.

Seager also points to another argument in favor of physicalism - the phenomenon of "intrusions from below" (Seager, 2014, 151). We find in nature what is evidently a hierarchy of structure, which sustain autonomous descriptions in terms of their own laws. However, we find that the autonomy of a given level in the natural hierarchy is broken from time to time, in that the laws of that level fail to hold because of some effect of a lower level. For example, in biology population dynamic equations are mathematically continuous, while any real biological population is composed of discrete individuals. This leads to anomalies called lattice effects, which are intrusions from below, interfering, so to speak, with the representation of the population as a continuous variable. The obvious explanation for the prevalence of intrusions from below is that the hierarchy of levels in nature exemplifies a system of ontological dependence, exactly as predicted by physicalism.

In spite of this impressive run of success of physicalist theories there is one particular phenomenon – consciousness – that has so far resisted inclusion into the physicalist worldview. I will return to this in Section 3.6. But first in Section 3.4 I examine in more detail the nature and scope of science of physics itself, and in Section 3.5 I consider what the content of modern physics tells us about reality.

3.4 PHYSICS AND MATHEMATICS

Physics is generally regarded as the most fundamental science. James Ladyman and Don Ross formulate this idea as the Primacy of Physics Constraint -

> "Special science hypotheses that conflict with fundamental physics, or such consensus as there is in fundamental physics should be rejected for that reason alone. Fundamental physical hypotheses are not symmetrically hostage to the conclusions of the special sciences" (Ladyman and Ross, 2007, 45).

Their argument for this constraint is based on two complementary inductive arguments from the history of science, which mirror Seager's optimistic induction within physics itself. Firstly they argue that in the history of science a succession of specific hypotheses to the effect that irreducibly non-physical entities and processes fix the chances of physical outcomes have failed. And secondly, a succession of processes in living systems, and notably in the parts of living systems dedicated to cognition, have come to be largely or entirely understood in physical terms.

Ladyman and Ross stress however, that the kind of fundamental physics they are talking about is not the classical physics of objects, collisions, and forces. Rather they argue that contemporary physics motivates a metaphysics of *ontic structural realism.* On this view, fundamental physics "denotes a set of mathematically specified structures without self-individuating objects, where any measurement taken anywhere in the universe is in part measurement of these structures" (Ladyman and Ross, 2007, 44). This metaphysics argues for the reality of a skeletal world of relations without relata that we are assured lies behind mere appearances, a world that is simpler than the world of our everyday life.

In identifying *mathematically* specified structures at the heart of fundamental physics, Ladyman and Ross are following in the tradition of Galileo who is best known for laying the foundations for the scientific revolution by identifying mathematics as the language of natural philosophy. In *The Assayer* he writes –

> "Philosophy [by which he means 'natural philosophy', which became physical science], is written in this grand book, the universe, which stands continually open to our gaze, but it cannot be understood unless one first learns to comprehend the language and read the letters in which it is composed. It is written in the language of mathematics, and its characters are triangles, circles, and other geometrical figures, without which it is humanly impossible to understand a single word of it; without these, one wanders about in a dark labyrinth" (Galileo, 1623, 237-238).

However, Philip Goff (2017) draws attention to the metaphysical picture of reality that Galileo was working with. Before Galileo, philosophers took the world to be full of sensory qualities such as colors, smells, tastes and sounds. Galileo got around the problem of mathematically dealing with these qualities by locating them in the soul or consciousness rather than in the world. As Galileo himself puts it -

> "… I think that tastes, odours and colours, and so on are no more than mere names as far as the object in which we place them is concerned, and that they reside only in the consciousness. Hence, if the living creature were removed, all these qualities would be wiped away and annihilated" (Galileo, 1623, 274-7).

Thus, for methodological reasons, Galileo placed the sensory qualities we encounter in conscious experience outside the material domain studied by physics. The subsequent success of physics in enabling the manipulation of the material world in all sorts of ways, led some scientists and philosophers to think that physics is on its way to giving us a complete account of fundamental reality – a theory of everything.

But Goff argues that this is the wrong conclusion to draw –

> "The success of mathematical physics resulted from limiting the scope of enquiry; by putting the sensory qualities we encounter in conscious experience - colours, smells, tastes, sounds - outside of the domain of the physical sciences, we are able to give a purely mathematical description of what's left over. But those qualities that Galileo took out of the material world still exist somewhere and must still be accounted for somehow" (Goff, 2017, 11).

Thomas Nagel makes a similar point-

> "The great advances in the physical and biological sciences were made possible by excluding the mind from the physical world. This has permitted a quantitative understanding of the world, expressed in timeless, mathematically formulated physical laws. But at some point

> it will be necessary to make a new start on a more comprehensive understanding that includes the mind" (Nagel, 2012, 8).

Goff argues that the fact that physical science has done well since sensory qualities were put outside its domain of enquiry, gives us no grounds for thinking it has the potential to adequately deal with sensory qualities themselves, because there is more to reality than can be captured in the austere mathematical vocabulary of physics.

As John Myhill (1952) points out not everything in the world is 'computable' or 'listable'; the world also has 'prospective' features. He identifies these features by extending the lessons learnt from the theorems of Gödel (1934), Church (1936), and Turing (1936) about the scope and limitations of logical systems. The quantifiable aspects of the world have the property of being computable, in that there exists a definite procedure for deciding if any candidate does or does not possess the required property. However, Turing and Church's restrictions (expressed in terms of unsolvability) mean that not every property of the world is computable. Listable properties are more elusive in that we can construct a procedure, which will list all the entities, which possess the required property, but there is no way of systematically generating entities which do not possess the required property. Most logical systems are listable but not computable in that all their theorems can be listed but there is no automatic procedure for inspecting a statement and deciding whether it is or is not a theorem. Gödel's theorem (expressed in terms of undecidability) shows that not every property of logical systems is listable. As I mentioned in the Introduction, Gödel's theorem shows that no formal system can be *both* consistent and complete because any formal system can be shown to contain at least one statement that is undecidable within the system. Myhill (1952) claims that the fact that we can know the truth of such a statement shows that there are features of the world such as truth, beauty and simplicity, which are not computable or listable; he calls them 'prospective' features. Although these prospective features cannot be recognized or generated by any series or

sequence of logical steps, we find ourselves habitually using them. We know them from the way we find ourselves living.

The existence of these prospective features leads John Barrow (2007) to conclude that any purely mathematical or formal 'theory of everything' can never be sufficient to reveal everything about our universe. It is not even sufficient to reveal everything about the physical world. As Bertrand Russell pointed out – "Physics is mathematical not because we know so much about the physical world, but because we know so little: it is only its mathematical properties that we can discover. For the rest, our knowledge is negative" (Russell, 1927, 125). Physics gives us knowledge of structure without giving us any understanding of the nature of what has that structure.

Thus Barrow argues that while a mathematical understanding is a necessary part of a full understanding of things, "no non-poetic account of reality can be complete" (Barrow, 2007, 265). And he points out that while a purely mathematical or formal 'theory of everything' cannot predict these prospective aspects of reality; "yet, strangely, many of these qualities will themselves be employed in the human selection and approval of an aesthetically acceptable Theory of Everything" (Barrow, 2007, 246). Any theory whatsoever is a human creation: 'everything' has to encompass the human features that make the creation of theories possible.

Bryan Van Norden (2017) makes a similar argument to claim that natural science will never replace philosophy because sciences like physics limit their enquiry to particular aspects of reality using particular methods and any claim that there is nothing beyond physics cannot be provided just from within physics -

> "Physics uses a particular methodology, M, to study reality insofar as it is physical, P. But the question we are asking is whether there is a reality that is not-P. Since M is the methodology we use to study reality only insofar as it is P, we cannot use that same methodology to explore whether there is anything that is not-P."

Van Norden's point is that you have to straddle a limit, conceptually speaking, in order to define it. Therefore you cannot show from within the

limits of something that there is nothing outside that limit. Thus he argues that –

> "… if you try to show that there is nothing beyond the factual and methodological limits of physics, you have already transcended the limits of physics." (Van Norden, 2017, 164).

Also, George Ellis argues that even within its own domain physics is self-referentially incomplete –

> "there is no physics theory or experiment that can determine what will be the next experiment to be undertaken by the experimenter or theory to be created by the theorist. If physics were causally complete, this would not be the case" (Ellis, 2005, 6.1).

Ellis is pointing out that the subject matter of physics has nothing to say about the way that physics as a scientific discipline operates. Ellis argues that physics makes possible, but does not causally determine the higher-order layers of structure and meaning. Thus it cannot replace psychology, sociology, politics, and economics as autonomous subjects of study. On this view, the higher levels in the hierarchy of complexity have real autonomous causal powers, largely independent of the lower levels and indeed controlling their context and hence their outcomes.

These arguments - that physics has a deliberately limited domain of enquiry and is self-referentially incomplete - suggest that physics, on its own, will not give us a metaphysically complete picture of reality. And, in the next section, I argue that the evidence from within physics itself does not give any certainty of even a purely physical account of reality.

3.5 THE INTERPRETATION OF QUANTUM PHYSICS

In the early twentieth century, with the development of quantum mechanics, physics changed radically. It became impossible to explain in intuitive everyday terms, and there was no clear agreed interpretation of

what it meant for how we understand reality. A century later, because of its success in predicting how matter behaves, quantum mechanics is still the dominant paradigm; but the interpretation of what it means has still not been settled. As the astronomer Adam Frank (2017) puts it –

> "after more than a century of profound explorations into the subatomic world, our best theory for *how matter behaves* still tells us very little about *what matter is*. Materialists appeal to physics to explain the mind, but in modern physics the particles that make up a brain remain, in many ways, as mysterious as consciousness itself."

When Albert Einstein and Max Planck introduced the idea of the quantum at the beginning of the 20th century, they swept away the old classical view of reality, and we have never managed to come up with a definitive new reality to take its place. There is, as yet, no agreed interpretation of what quantum physics tells us about reality.

Prior to the 20th century, the location and velocity of a particle of matter were understood in terms of Newton's equation $F = ma$. Apply a force F to a particle of mass m and it moves with acceleration a. However, early in the 20th century, as scientists examined matter at the atomic and sub-atomic level, a more complicated and fuzzy picture emerged. At this level matter appears to exhibit strange behavior described by a form of mathematics called quantum mechanics. The Austrian physicist Erwin Schrödinger (1926) developed an equation for dealing with these quantum objects. Like Newton's $F = ma$, Schrödinger's equation is the mathematical machinery for doing physics; it describes how the state of a particle is changing. But while Newton's equation gave exact answers for position and velocity, Schrödinger's equation only gives answers in terms of a wave function with probabilities of different positions and velocities. The particle appears to have many positions and many velocities.

One influential interpretation of wave functions was proposed by the Danish physicist Niels Bohr (1934) and became known as the Copenhagen Interpretation. This says that the wave function is a superposition of all possible states of the particle, and that this collapses to one particular state

when an observer makes a measurement. This, then, gives a critical role to the interaction between the observer and the wave function.

According to Bohr, it is meaningless to speak of the properties of atoms in-and-of-themselves. He argued that the purpose of science was not to reveal "the real essence of the phenomena" but only to find "relations between the manifold aspects of our experience". On this view, quantum mechanics is a theory that speaks only to our knowledge of the world. The measurement problem associated with the Schrödinger equation highlighted this barrier between epistemology and ontology by making explicit the role of the observer (that is: us) in gaining knowledge.

Other interpretations want to retain the ideal of access to an objective world. In a lecture in Dublin in 1952, Schrödinger stated that when his equation seems to be describing several different histories, they are 'not alternatives but really all happen simultaneously'. Just before he said this he warned his audience that what he was about to say might "seem lunatic", and it seems that he never pursued the idea any further (Deutsch, 2011, 310).

However, a few years later Hugh Everett (1955), independently, proposed a many-worlds (multiverse) interpretation of quantum theory in which the authority of the wave function and its governing Schrödinger equation was taken as absolute. Measurements didn't suspend the equation or collapse the wave function; rather, they made the Universe split off into many (perhaps infinite) parallel versions of itself. Thus, for every experimenter who measures an electron *over here,* a parallel universe is created in which her parallel copy finds the electron *over there.* This interpretation maintains a materialist view but it comes with a steep price. It keeps the reality in mathematical physics but the price you pay for this position is an infinite number of parallel universes that are infinitely splitting off into an infinity of other parallel universes - infinitely. Some versions of multiverse theories involve a special role for the observer or measurer in that it is the act of measurement that make universes split into parallel versions.

David Deutsch (2011) is a strong advocate of the multiverse theory. He understands science as producing explanations – "an account of what is

really out there" (ibid. 263). He argues that quantum theory is "the deepest explanation known to science" (ibid. 262), and taking quantum theory seriously involves understanding reality as a multiverse. His account involves the idea that entities can be fungible, by which he means identical in literally every way except that there are two of them. He acknowledges that this is contrary to Leibniz's doctrine of 'the identity of indiscernibles' (Leibniz, 1686, section 9): Deutsch dismisses this saying - "he was mistaken" (Deutsch, 2011, 266). Deutsch argues that although it violates many assumptions of common sense and of all previous science, "this seemingly alien territory is the reality of which we and everything we experience are part. There is no other" (ibid. 262). He argues that although our universe approximately obeys the laws of classical (pre-quantum) physics, we know of the rest of the multiverse, and can test the laws of quantum physics, because of the phenomenon of quantum interference (ibid. 304).

David Bohm's (1980) theory of a multidimensional reality was also an attempt to integrate the novel features of quantum theory into a theory of "the totality of all there is". His central idea is the unbroken wholeness of the totality of existence as an undivided flowing movement (*holomovement*) without borders. And, according to his concept of implicate order, the totality of existence is enfolded within each region of space and time (ibid.172). He developed this theory specifically to accommodate both the theory of relativity and quantum theory, which he argued directly contradict each other, but which both, in different ways, point to the concept of wholeness (ibid.176).

An alternative interpretation sees physics as no longer a description of the world in-and-of itself; instead, it's a description of the rules for our interaction with the world. On this view quantum mechanics is a calculus – a machine for predicting experimental results.

According to Quantum Bayesianism or QBism, proposed by Christopher Fuchs and Rudiger Schack (2013), the irreducible probabilities in quantum mechanics tell us that it's really a theory about making bets on the world's behavior (via our measurements) and then updating our knowledge after those measurements are done. This interpretation can be described as

'participatory realism', because human agency is woven into the process of doing physics as a means of gaining knowledge about the world. Christopher Fuchs stresses however, that the label 'participatory realism' emphasizes rather than relinquishes the idea of reality; what it claims is that reality is more than any third-person perspective can capture. As he argues –

> "… far from instances of instrumentalism or antirealism, these views of quantum theory should be regarded as attempts to make a deep statement about the nature of reality" (Fuchs, 2016, 113).

On this view then, the equations of quantum physics don't refer just to the observed atom but also to the observer and the atom taken as a whole in a kind of 'observer-participancy'. QBism points explicitly to our failure to include the observing subject that lies at the root of quantum weirdness. As David Mermin (2014) puts it - 'QBism attributes the muddle at the foundations of quantum mechanics to our unacknowledged removal of the scientist from the science.' But as Frank (2017) points out putting the perceiving subject back into physics would seem to undermine the whole materialist perspective – "A theory of mind that depends on matter that depends on mind could not yield the solid ground so many materialists yearn for."

Frank (2017) points out that as yet there is no way to experimentally distinguish between these widely varying interpretations. Which one you choose is mainly a matter of philosophical temperament. But he does question why certain intuitively implausible quantum interpretations are widely preferred over others within the research community – "Why does the infinity of parallel universes in the many-worlds interpretation get associated with the sober, hard-nosed position, while including the perceiving subject gets condemned as crossing over to the shores of anti-science at best, or mysticism at worst?"

Thus, he argues that the high ground of materialism deflates when followed to its quantum mechanical roots, because it then demands the acceptance of metaphysical possibilities that seem no more 'reasonable' than other alternatives – "We know that matter remains mysterious just as mind remains mysterious, and we don't know what the connections

between those mysteries should be. Classifying consciousness as a material problem is tantamount to saying that consciousness, too, remains fundamentally unexplained."

He concludes that "the unresolved democracy of quantum interpretations means that our current understanding of matter alone is unlikely to explain the nature of mind. It seems just as likely that the opposite will be the case."

So, at the present time, physics does not give us an agreed picture of reality. A central problem in all the theories is how to account for the observer or consciousness. In the next section I examine how consciousness has become the critical problem not just for physics, but also for a physicalist metaphysics.

3.6 THE BARRIER OF CONSCIOUSNESS AND THE BURDEN OF PROOF

In spite of the great progress that science has made in creating the scientific picture of the world, there has been a conspicuous lack of even a hint of how consciousness could be explained in terms of the physical. Jaegwon Kim argues that although "(w)e can imagine designing and constructing novel physical systems that will instantiate certain cognitive capacities and functions (e.g. perception, information processing, inference and reasoning, and using information to guide behavior) … it is difficult to imagine our designing novel devices and structures that will have phenomenal experiences; I don't think we have any idea where to begin" (Kim, 2010, 14). As Seager puts it – "the forward march of physicalism has been impeded, if not halted by the phenomenon of consciousness" (Seager, 2014, 9).

As I noted in Section 3.4, Philip Goff (2017) argues that from the start of science, consciousness was excluded from its domain of study. And philosophers have remarked on the difficulties in knowing how it could be studied by science. In his paper *What Is It Like to Be a Bat?* Thomas Nagel (1974) argues that something is conscious if *there is something it is like* to be

that thing. He explains that to be conscious is to have subjectivity – to experience the world from an individual point of view. Thus, as he puts it, we might be able to imagine what it would be like for us to do some of the things that we observe bats doing, but it seems very hard to see how we could use our imagination to know what it is like *for a bat*, to do those things. And if consciousness is intrinsically perspectival or subjective, it seems difficult to understand how it could be investigated by science, which involves a non-perspectival objective account. In his early writings, Nagel is not entirely pessimistic that science might someday explain consciousness as a physical process. He thought that it might be possible to develop concepts that would make this understandable. But in his 2012 book, *Mind & Cosmos*, he argues that reductive materialism is a world view that is ripe for displacement –

> "What has to be explained is not just the lacing of organic life with a tincture of qualia but the coming into existence of subjective individual points of view – a type of existence logically distinct from anything describable by the physical sciences alone" (Nagel, 2012, 44).

Frank Jackson's 'Mary argument' (Jackson, 1982) makes a similar point. Mary is a physically omniscient neuroscientist who has always lived in a black and white environment. The thought experiment works on our intuition that Mary would not know what it is like to see red prior to her first experience of it. It seems convincing because there seems to be no way to begin to understand how consciousness could epistemologically depend on brain activity. Also, David Chalmers (1996) argues for a modal gap between the physical and consciousness because of the obvious absence of any sense of how consciousness could be reductively epistemologically dependent on the physical.

These arguments work by arguing that a failure to demonstrate how epistemological dependence is possible leads to doubts about physicalism. Although physicalism is primarily an ontological theory, the source of evidence in favor of ontological dependence is the exhibition of reductive epistemological dependence. Thus, Seager argues that the persistent failure

to discover epistemological dependence suggests a failure of ontological dependence –

> "The strength of physicalism is its long history of success, but this is a history of revelation of patterns of epistemological dependence. Thus, somewhat paradoxically – in the face of the stubborn reluctance of consciousness to be integrated into the scientific metaphysics via the normal route of epistemological dependence, the long history of success of physicalism is evidence *against* it" (Seager, 2014, 12).

Seager also points to a paradox in the scientific picture of the world. According to the scientific picture of the world all emergence is conservative, so all emergent features are epiphenomenal – that is such features cannot possess distinctive causal efficacy of their own. Seager understands emergent features as being what Dan Dennett (1991) calls 'real patterns'. Patterns are structures, and relations among structures, that are visible from certain viewpoints – "any relation among data" (Ladyman and Ross, 2007, 228). Ladyman and Ross (2007) endorse Dennett's 'patterns' to develop a scale relative ontology that they term 'rainforest realism'. They make a distinction between "mere patterns" – "stable relationships in data" – and "real patterns" by means of mathematical information theory –

> "A [mere] pattern is redundant, and not an ultimately sound object of scientific generalization or naturalized ontology, if it is generated by a [real] pattern of greater computational power (lower logical depth)" (Ladyman and Ross, 2013, 108).

Their claim is that "what (really, mind-independently) exists should be relativized to (real, mind-independent) scales at which nature is measurable" (Ladyman and Ross, 2007, 200), and what exists are just these scale relative patterns. As they put it – "To be is to be a real pattern" (Ladyman and Ross, 2007, 226); and "it's real patterns all the way down" (ibid. 228). They state their metaphysical thesis as - "the world is the totality of non-redundant statistics (not things)" (Ladyman and Ross, 2013, 111).

Seager (2012) points to a difficulty with this that arises from the peculiar nature of patterns -

- "… patterns are somewhat odd. They inhabit a curious zone midway between, as it were, objectivity and subjectivity, for patterns are there to be seen but have no function if they are not seen. By the former, I mean that patterns are not just in the eye of the beholder; they are really in the world (it is not optional for us to decide that salt does or does not dissolve in water or that a checkmate is forced in two moves) and they provide us with an indispensably powerful explanatory and predictive grip on the world. By the latter, I mean that the only role they have in the world is to help organize the experience of those conscious beings who invent concepts for them and then think in terms of those concepts" (Seager, 2012, 166-7).

Thus, he argues that –

"although the world is rightly described as exemplifying a host of patterns, the world itself, so to speak, has no use for them. … it is the fundamental physical features of the world which organize the world into all the patterns it exemplifies and they do this all by themselves, with no help from 'top-down' causation" (Seager, 2012, pp. 166-7).

Because the metaphysically real features, which drive the world from state to state, all reside at the fundamental level alone, this generalized epiphenomenalism leads to the *paradox of consciousness*. This arises because the conservative emergents (Seager's 'natural fabrications' or Dennett's, and Ladyman and Ross's 'real patterns') have no actual role in the world, apart from being identified and used by conscious entities that find them, or concepts of them, helpful in categorizing and organizing their experience. As Seager puts it - "[t]hese emergents stand as mere epistemic potentials, awaiting take-up by some appreciative mind". But this leads to a paradox -

"The paradox looms when we realize that consciousness itself, according to the scientific picture of the world, must be merely a conservatively emergent feature. The scientific picture strongly suggests that consciousness is the product of vastly complicated neural systems, which appeared in the universe long after its creation. But if consciousness itself is a conservatively emergent feature then it too

> stands as epiphenomenal and a mere epistemic resource that appears only to … consciousness" (Seager, 2012).

For Ladyman and Ross, the paradox looms when it is realized that if 'all there is' is real patterns, then the observer of these real patterns must also be a real pattern.

So, it appears that the scientific picture of the world is incoherent. It regards the mind as a high-level feature of reality, but then it finds that a conscious mind is needed to make sense of high-level features.

3.7 ESCAPE ROUTES

Consciousness appears to create a barrier to a complete physicalist picture of reality because including it within the scientific picture of the world leads to a paradox.

Seager (2012, 2014) investigates four possible alliteratively termed routes to escape from this unsatisfactory position – watchful waiting, embrace emergence, favour fundamentality or modify metaphysics. I will outline the first three options in the remainder of this chapter and consider the last option in chapter 5.

3.7.1 WATCHFUL WAITING

This option insists that consciousness is a standard conservatively emergent phenomenon that will eventually be fitted into the scientific picture of the world in the usual way. Most scientists would say that current science is not complete and that there is a lot more to be investigated. So the fact that no comprehensive theory of consciousness has yet developed shouldn't be a particular worry. They can argue that given more time and more money the problem will be solved. And they can point to some progress already made. A lot of work has been done on identifying the neuronal correlates of consciousness – linking specific aspects of consciousness to particular neuronal mechanisms or activities. Also, by understanding consciousness in terms of information theory, Giulio Tononi

(2004, 2008) claims to have developed a precise measure, called phi, which captures the extent of consciousness by measuring the degree to which a system's information is integrated.

However, to understand consciousness it looks like we are going to need more than just more correlational data and measurement. We need a fundamental change in our understanding of the physical world, consciousness, or both, which will permit new, but recognizably legitimate, mechanisms of conservative emergence. As I mentioned in the previous section, Thomas Nagel (1974) in his early writings thought that perhaps sometime in the future we could develop concepts that would allow us to understand how consciousness could be a physical process. Maybe we can't now see how this would be possible, but in the twelfth century it would have not have been possible to understand how matter and energy could be interchangeable. Similarly, Roger Penrose (1989) in *The Emperor's New Mind,* argues that consciousness can only be scientifically understood via a radical change in physics, which will do no less than introduce some features that transcend Turing computability. The reason Penrose offers for this is that he thinks that Gödel's incompleteness result entails that the intellectual power of the conscious mind somehow goes beyond the power of algorithmic computation. The connection to consciousness arises because the problem appears only in the case of conscious thoughts in which the Gödel problem is appreciated.

The physicist Max Tegmark (2015, 2017) is an example of someone who thinks that the 'watchful waiting' strategy will work. He claims to give an account of how it might be possible to develop concepts that allow us to understand consciousness as a physical process. He argues that consciousness is best understood as a state of matter, just like a solid, liquid or gas. Just as the key difference between a solid, a liquid and a gas lies not in the type of atoms, but in their arrangement, so consciousness is a state whose complex behavior emerges from many simple interactions. Like waves, which are substrate-independent in the sense that they can occur in all liquids, regardless of their atomic composition, consciousness is to be understood at the substrate-independent level of information. Tegmark

contends that "consciousness is the way information feels when processed in certain complex ways" (Tegmark, 2017, 301). He doesn't claim that this yet gives us a full understanding of 'the way information feels', but he does propose certain principles that are necessary but not sufficient conditions for a conscious system –

- the *information principle* – it must have substantial information capacity.
- - the *dynamics principle* – it must have substantial information-processing capacity, and it is this processing rather than the static information that must be integrated.
- - the *independence principle* – it must have substantial independence from the rest of the world.
- - the *integration principle* – it cannot consist of nearly independent parts.

Temark argues that the first three principles imply *autonomy* - the system is able to retain and process information without much outside interference, hence is able to determine its own future. The four principles together mean that a system is autonomous but its parts aren't (Tegmark, 2017, 304/5).

Tegmark's assessment of his attempt to ground consciousness in fundamental physics is that "much work remains and the jury is still out on whether it will succeed" (Tegmark, 2015, 97). We must watch and wait for a bit longer.

But Seager (2012,192) argues that no matter what new science or conceptual revolution occurs, the nature and role of conservative emergents remains clear. They are casually inefficacious and exist solely as potential epistemic resources for conscious beings intent upon understanding, explaining and predicting the world. Substrate-independent phenomena like 'waves' and 'information' seem like good candidates for potential epistemic resources for conscious beings. But if consciousness is grounded in information, then, consciousness itself is simply another conservative emergent, and this leads to incoherence.

Tegmark, himself, does realize the importance of accounting for conscious observers. He notes that in Einstein's theory of general relativity,

we model the 'observer' as a fictitious disembodied massless entity having no effect whatsoever on what is observed, whereas the textbook interpretation of quantum mechanics states that the observer does affect the observed. Also understanding the observer seems to be central to other fundamental physics mysteries such as why our universe appears so orderly, and why time appears to flow in a forward direction. He concludes – "If we can figure out how to identify conscious observers in any physical system and calculate how they will perceive their world, then this might answer these vexing questions" (Tegmark, 2015, 97).

The difficulty with this is that 'how they will perceive their world' does not seem to be the kind of thing that could be calculated. As Michel Bitbol puts it –

> "Scientists who believe that solving many such 'easy problems' about consciousness will finally clear up the harder problem of its physical origin, look like somebody who believes one can finally reach the horizon by walking far enough" (Bitbol, 2008, 10).

Waiting for an explanation of how this might be possible may not make sense. Considerations like this have led some philosophers to consider more radical explanations of consciousness.

3.7.2 EMBRACE EMERGENCE

According to Seager's scientific picture, the world has many conservatively emergent features. These are features that have no causal power in themselves but they are useful to conscious entities in their understanding of the world. On this view then the conscious entities themselves cannot be just other conservatively emergent features. This suggests that these conscious entities must be emergent in a stronger or more radical sense such that they have some novel causal power of their own, over and above the powers of their physical constituents. This is consistent with the views of the British emergentists such as John Stuart Mill (1843), Samuel Alexander (1920), C.L. Morgan (1923), and C.D. Broad (1925) who claimed that radical emergence is a widespread feature of a

hierarchical structure of reality. Many of the examples that these emergentists gave have now been shown, by scientific developments, to be merely conservatively emergent features. But consciousness remains to be fully explained. As Brian P. McLaughlin (1992, 91) puts it – "Consciousness, it seems, is the last refuge of an Emergentist".

Seager explains how radical emergence is claimed to occur in the case of consciousness -

> " … when certain physical systems attain sufficient and appropriate complexity, as in the human brain, there emerges a novel feature, consciousness, with all its attendant phenomenal qualities. This new property of the brain is lawfully related to its subvening brain state but is not determined solely by the underlying physical state and purely physical law. It has its own causal powers which can act back on the physical world. This is *genuine* downward causation, not the anodyne simulacrum allowed by conservative emergence. There is no sense in which this new feature is a mere epistemic potential or pattern – it has its own stance-independent being" (Seager, 2012, p.197).

According to radical emergence then, there are irreducible laws of emergence (what C.D. Broad (1925) terms 'trans-ordinal laws') that go beyond the laws of fundamental physics. Everything may still have exclusively physical constituents, but in certain situations, the way these physical things combine may give rise to features that are not determined by the physical constituents and purely physical laws. As C.D.Broad puts it –

> "[T]he characteristic behaviour of the whole *could* not, even in theory, be deduced from the most complete knowledge of the behaviour of its components, taken separately or in other combinations, and of their proportions and arrangements in this whole" (Broad, 1925, 59).

Broad presents this as an ontological claim about the world: it's about "the differences of behaviour between different things" (Broad, 1925, 58), and it applies "even in theory". This amounts to a rejection of physicalism as a complete theory of theory of everything. The claim is that there are some other principles of organization apparent in the world. As Broad put it –

> ‘‘On the emergent theory we have to reconcile ourselves to much less unity in the external world and a much less intimate connexion between the various sciences’’ (Broad, 1925, 77).

On this view then, we would have to give up the hope of ever achieving the sort of unified scientific of the world that Barrow says most scientists are aiming for – “a monumental ‘Theory of Everything’ which will unite the laws of Nature into a single statement that reveals the inevitability of everything that was, is and is to come in the physical world” (Barrow, 2007, Preface). But Broad argues that there is nothing mysterious or unscientific about his account -

> ‘‘There is nothing, so far as I can see, mysterious or unscientific about a trans-ordinal law or about the notion of ultimate characteristics of a given order. A trans-ordinal law is as good a law as any other; and, once it has been discovered, it can be used like any other to suggest experiments, to make predictions, and to give us practical control over external objects. The only peculiarity of it is that we must wait till we meet with an actual instance of an object of the higher order before we can discover such a law; and that we cannot possibly deduce it beforehand from any combination of laws which we have discovered by observing aggregates of a lower order’’ (Broad, 1925, 79).

As Joel Walmsley points out this still allows for “a set of law-like regularities that can be used as the basis for a science of higher-level phenomena” (Walmsley, 2015, 195). So we can do science, but our science will never give us a complete theory of everything because we cannot know when we will come across a new emergent feature. We can have a current scientific picture of the world, but it will involve brute facts and will therefore have limited usefulness in producing an overall metaphysical picture.

Radical emergence can be seen as the minimum addition to physicalism that is necessary to account for phenomena such as consciousness, which do not fit into the purely physical account. Substance dualists, notably Descartes, accounted for these phenomena as separate elements added to the physical brain from the outside. Radical emergentists, on the other hand,

want to account for them as law abiding products of the physical brain. Here is how William Hasker (1999) accounts for mental properties -

> "… mental properties are 'emergent' in the following sense: they are properties that manifest themselves when the appropriate material constituents are placed in special, highly complex relationships, but these properties are not observable in simpler configurations nor are they derivable from the laws which describe the properties of matter as it behaves in these simpler configurations" (Hasker, 1999, 189-190).

On this view, emergentism involves properties of matter that reveal novel causal powers that are not evident in the absence of consciousness.

Although radical emergence is often thought of only in terms of properties, Hasker argues that while this is necessary, it is not sufficient to capture the kind of being which humans have. He argues that the experience of a unified consciousness cannot be explained in terms of any combination of properties of, and relations between, the material constituents of the brain –

> " A conscious experience simply *is* a unity, and to decompose it into a collection of separate parts is to falsify it. So it is not enough to say that there are emergent properties here; what is needed is an *emergent individual*, a new individual entity which comes into existence as a result of a certain functional configuration of the material constituents of the brain and nervous system" (Hasker, 1999, 190).

According to this even more radical emergentist theory – Hasker calls it 'emergent.2' – its not just properties that are emergent, but also individuals.

Radical emergence does seem to be a coherent theory. Maybe that's just the way the world is and we need an emergentist theory to describe it. But radical emergence can seem like an 'ad hoc' addition to the scientific picture of the world introduced solely to explain consciousness. As C.D. Broad explains it, trans-ordinal laws are discoverable brute facts that "must simply be swallowed whole with that philosophic jam which Professor Alexander calls 'natural piety'" (Broad, 1925, 55). Samuel Alexander himself takes the term 'natural piety' from William Wordsworth's poem *My Heart Leaps Up*, in

which the poet uses it to describe his experience of a rainbow. Alexander uses it to express the fact that emergent qualities are inexplicable –

> "(t)he existence of emergent qualities … is something to be noted, as some would say, under the compulsion of brute empirical fact, or, as I would prefer to say in less harsh terms, to be accepted with the "natural piety" of the investigator. It admits no explanation" (Alexander, 1920/1979, 46).

While brute facts accepted with natural piety can form part of our current scientific picture of the world, such a picture could not claim to be consistent with physicalism. At best, in Jaegon Kim's (2005) terms it would be "a slightly defective physicalism – physicalism manqué but not by much". However, physicalism is a universal claim that everything is fundamentaly physical and therefore can be explained with physical terms. But brute facts are facts for which there is no explanation. Having no explanation for consciousness, which is the source of all our knowledge, is a serious not a slight defect – near enough is not good enough.

In fairness to Hasker, he does attempt to give an account of how his emergent individuals might be understood in terms of a physical universe. He presents his concept of higher emergence using an analogy with the various 'fields' that are familiar from physical science, such as, the magnetic field or the gravitational field. A magnetic field is a real, existing, concrete entity, distinct from the magnet, which produces it. Once generated, the field exerts a causality of its own, on the magnet itself as well as on other objects in the vicinity. Thus Hasker argues –

> "as a magnet generates its magnetic field, so the brain generates its field of consciousness. The mind, like the magnetic field, comes into existence when the constituents of its 'material base' are arranged in a suitable way – in this case, in the extremely complex arrangement found in the nervous system of humans and other animals. And like the magnetic field, it exerts a causality of its own; certainly on the brain itself, and conceivably also on other minds (telepathy) or on other aspects of the material world (telekinesis)" (Hasker, 1999, 190-191).

Hasker warns that this is merely an analogy; magnetic fields presumably can be understood in conventionally physical terms. However it is meant to suggest how there could be a close natural connection between mind and brain, as contrasted with the disparity between mind and matter postulated by Cartesianism. Hasker argues however that, in order to be realist about the phenomenon of mind itself, we have to recognize that a great many mental process are *irreducibly* teleological, and cannot be explained by or be supervenient upon, brain processes that have a complete mechanistic explanation. Thus he calls his position – emergent dualism.

Interestingly, Hasker recognizes that his theory has consequences for how we understand the power of matter –

> "… when suitably configured, it [matter] generates a field of consciousness that is able to function teleologically and to exercise libertarian free will, and the field of consciousness in turn modifies and directs the functioning of the physical brain" (Hasker, 1999, 195).

Thus Hasker's consideration of the implications of his theory of emergence leads him to consider the possibility that all matter is potentially conscious. This kind of consideration is the focus of Seager's third escape route.

3.7.3 FAVOUR FUNDAMENTALITY

The option of favouring fundamentality involves accepting that consciousness is in some way a fundamental irreducible feature of the world. There are different ways in which this has been developed including neutral monism, dual aspect theory and panpsychism.

Neutral monism, championed by Bertrand Russell, holds that there is an underlying neutral foundation of reality to which both mind and matter reduce in some sense but which itself is neither mental nor physical -

> "The stuff of which the world of our experience is composed is, in my belief, neither mind nor matter, but something more primitive than either. Both mind and matter seem to be composite, and the stuff of

which they are compounded lies in a sense between the two, in a sense above them both, like a common ancestor" (Russell, 1921, 7).

A difficulty with this view is that, contrary to Ockham's razor, this theory posits a third new kind of entity to explain the relationship between two known entities. As Seager puts it – "Neural monism posits an unknown something-or-other out of which emerges both matter and consciousness" (Seager, 2012, 199).

Dual aspect theory, on the other hand, regards mind and matter as two equally co-foundational aspects of an otherwise ineffable and inaccessible reality. According to this theory every element of fundamental reality is represented or expressed in both the mental and physical aspect. The pure form of dual aspect theory is a psycho-physical parallelism, such as that of Spinoza (1677/1996), in which mind and matter are isomorphic but utterly independent of each other. However, Seager argues that "dual aspect theory 'solves' the problem of consciousness by merely sidestepping around it" (Seager, 2012, 198).

Panpsychism is the view that all matter has a consciousness-involving nature. Philip Goff (2017) points out that there are two ways of developing the panpsychist position – micropsychism and cosmopsychism.

Micropsychism is the view that the smallest parts of the physical world have consciousness; or as William James (1890) put it, the smallest parts of the world are "mind-dust". In a section of *The Principles of Psychology* titled "Evolutionary Psychology Demands a Mind-Dust", he argues that "If evolution is to work smoothly, consciousness in some form must have been present at the very origin of things … Each atom of the nebula … must have had an aboriginal atom of consciousness linked with it."

Thomas Nagel (1979) lays out an argument for (micro)pansychism that accepts the scientific picture of the world almost completely, with the one proviso that science has missed some of the basic properties, which characterize its own fundamental entities. Nagel's argument goes as follows
- the only sort of emergence which is possible is conservative emergence,
- but it is impossible that consciousness should be a conservatively emergent feature of the world,

- yet mental properties are properties of organisms which are physical entities,
- thus 'if the mental properties of an organism are not implied by any physical properties but must derive from properties of the organism's constituents, then those constituents must have nonphysical properties from which the appearance of mental properties follows when the combination is of the right kind' (Nagel, 1979, 185),
- thus the constituents' properties must be mental,
- furthermore, since we can 'build' an organism out of any kind of matter, *all* matter must possess these elementary mental properties.

Similarly, Galen Strawson (2006) argues that it is only by supposing that human and animal consciousness emerges from more basic forms of consciousness, that we have hope of avoiding the emergence of animal consciousness being a brute and inexplicable miracle.

However, Seager argues that panpsychism is simply emergentism in disguise. The worry is that even if we suppose that consciousness is a fundamental feature of the world we still have to explain how complex minds come into being and are related with the elemental mental features. The only way to put together a panpsychist theory on the model of the scientific picture of the world requires the admission of a combination function that looks suspiciously like radical emergence, which is directly contrary to that model. David Chalmers (2016) calls this the combination problem of panpsychism. Philip Goff identifies a paradigmatic form of this problem – what he terms "the *subject-summing problem*: the difficulty making sense of how 'little' subjects could unify together to make a 'big' subject" (Goff, 2016, 169). This problem is similar to the problem I discuss in section 6.4 which questions how individual subjects could be unified into a group subject. Chalmers (2016) also identifies two other form of the combination problem: the *palette problem* relating to quality combination, and the *grain problem* relating to the combination of structure.

Considerations like this have led a number of scientists and philosophers of science to argue that this kind of 'bottom-up' picture of the universe is out-dated, and that contemporary physics suggests that in fact

we live in a 'top-down' – or 'holist' – universe, in which complex wholes are more fundamental than their parts. Jonathan Schaffer (2010) argues that the phenomenon of quantum entanglement is good evidence for holism. Entangled particles behave as a whole, even if they are separated by such large distances that it is impossible for any kind of signal to travel between them. According to Schaffer, we can make sense of this only if, in general, we are in a universe in which complex systems are more fundamental than their parts.

By combining holism with panpsychism, Itay Shani (2015) and Philip Goff (2016) argue for *cosmopsychism*: the view that the Universe is conscious, and that the consciousness of humans and animals is derived not from the consciousness of fundamental particles, but from the consciousness of the universe itself. I will discuss this view further in section 6.3. However, while it may solve some problems it does face a problem analogous to the combination problem – a subdivision problem - how are the individual conciousnesses that we experience derived from the unitary consciousness of the universe.

3.8 TOWARDS MODIFYING METAPHYSICS

The two options of radical emergence and favouring fundamentality seem to be inextricably linked. Without some account of consciousness at the fundamental level, radical emergence seems like a brute fact that we must accept with 'natural piety'. On the other hand, any account in terms of fundamental consciousness entities faces a problem of explaining how they combine (in the case of micropsychism) or subdivide (in the case of cosmopsychism) to constitute the kind of unified conscious subject that we know by being one. Seager says that both radical emergence and fundamentality suffer from the urge to minimize the disturbance consciousness introduces into the scientific picture of the world, yet both end up adding elements to the scientific picture for which there is no empirical evidence. As James Van Cleve (1990) put it the choice is between whether you add a bit of 'mind-dust' or 'magic'.

I do not claim to have shown that these problems, involving brute facts and combination or subdivision, are ultimately unresolvable. However, the difficulty of explaining them within an overall physicalist metaphysics justifies setting them aside to consider a more radical but simpler solution. This is Seager's fourth option – modifying metaphysics. It involves a rejection of scientific realism – the position holding that science provides the best, perhaps only, route towards knowledge of the ultimate nature of reality. This alternative view of science and nature need not undercut the explanatory power of science within its proper sphere. I examine this option in chapter 5, but first I examine one science that does deal directly with questions about what it is to be a human being.

CHAPTER 4. BIOLOGICAL SELVES & KNOWING-BY-BEING-ALIVE

4.1 Biological Individuals
4.2 Cells
4.3 Organisms
4.4 The Immune System
4.5 Autonomous Wholes
4.6 Hylomorphism
4.7 Knowing life by being alive

4.1 BIOLOGICAL INDIVIDUALS

As I argued in chapters 1 and 2, the way we find ourselves living is primarily as a subject of experience or self, and this subject of experience is constituted bodily. In chapter 3, I identified some problems that science in general has in accounting for selves and particularly the primary fact about the way we find ourselves living – consciousness, and also other aspects such as unity, independence, and agency, that seem to follow from it. Biology, as the study of living beings in general, is one science that includes an investigation of what it is to be a body or an organism, and therefore should have something to say about what it is to be self and to be conscious. As John Dupré put it –

> "Biology is surely the science that addresses much of what is of concern to us biological beings, and if it cannot serve as a paradigm for science, then science is a far less interesting undertaking than is generally supposed" (Dupré, 1993,1).

In this chapter I examine the ways in which the science of biology identifies individuals. I argue that many of the concepts used in biology are derived from our own sense of being alive. This reversal of the priority of self and science apparent in biology will be applied more generally to all science in the following chapter.

Biology, like any science, has to identify certain concepts or individuals that it picks out as significant objects of investigation – "what, in the living

world, constitutes a relatively well-delineated and cohesive unit" (Pradeu, 2016, 800) Answering this question is important in biology, for example, to be able to count offspring, to distinguish growth from reproduction, to make interspecies comparisons, to distinguish organisms from parts of organisms, and to determine if tissues from one living thing can be transplanted into another living thing. It is also important in understanding what we are.

One way of approaching this problem has been from the evolutionary perspective. From this perspective, individuals are identified as 'units of selection", characterized by variation, heritability, and differential fitness (Lewontin, 1970; Gould and Lloyd, 1999). Because these 'units of selection' can occur at different hierarchical levels, evolutionary individuals could be genes, chromosomes, organelles, or even populations, as well as organisms.

Another approach to identifying biological individuals is from the perspective of physiology – the "science of functions and constants of the functioning of living organisms" (Canguilhem, 1994). This involves investigating how distinct and heterogeneous components interact and constitute a cohesive whole, or regulated unit, which persists through time. This is the approach that I will follow.

4.2 CELLS

The fundamental component of living systems has long been identified as the cell. In 1858, the German biologist Rudolf Virchow wrote -

" … the cell constitutes the true organic unit, that is the ultimate irreducible form of every living element …" (Virchow, 1978, [1858], 23).

The components of the cell itself are entirely physical or chemical – but there is something about the organization of these components in a cell that allows the cell as unit to behave in a coherent way. Maturana and Varela (1973, 1980, 1987) describe the particular type organization that cells exhibit as autopoietic (self-producing). According to this theory, a cell produces its own components, which in turn produce it, in an on going circular process. As Evan Thompson (2007) explains it –

> "A cell is a thermodynamically open system, continually exchanging matter and energy with its environment. Some molecules are imported through the membrane and participate in the processes inside the cell, whereas other molecules are excreted as waste. Throughout this exchange, the cell produces a host of substances that both remain in the cell (thanks to the membrane) and participate in those very same production processes" (Thompson, 2007, 98).

Thompson thus sees the cell as paradigm case of an autonomous system, which arises from a process of dynamic co-emergence in which part and whole co-emerge and mutually specify each other (Thompson, 2007, 65). His description of the cell's independent yet interdependent existence is the forerunner of the kind of existence that we find ourselves living. As I put it in chapter 1, independence and interdependence are primary facts of our existence.

And Thompson points to "the deep continuity of life and mind", so that "the *self-organizing* features of mind are an enriched version of the self-organizing features of life" (Thompson, 2007, ix).

In a passage reminiscent of Dennett's (1991a) real patterns, Thompson describes a cell as a figure that emerges out of a chemical background –

> "a cell stands out of a molecular soup by creating the boundaries that set it apart from what it is not. Metabolic processes within the cell determine these boundaries, but the metabolic processes themselves are made possible by those very boundaries. In this way the cell emerges as a figure out of a chemical background" (Thompson, 2007, 99).

Varela (2000) himself proposed the following three-part definition of autopoiesis –

For a system to be autopoietic,

(i) the system must have a semipermeable boundary;

(ii) the boundary must be produced by a network of reactions that take place within the boundary; and

(iii) the network of reactions must include reactions that regulate the components of the system.

This operational closure of the autopoietic system does not mean that it becomes entirely independent from its environment. Rather the operational closure can only be sustained by energetic or material (thermodynamic) openness.

Biologists now have a good understanding about the functional organization of cells. But questions about when, where and how the first cell (or cells) came into existence, and whether this has only happen once in one place and time, or more, has yet to be determined. However, the significance of the appearance of cells seems apparent. Thompson (2007) argues that even at the minimum level of cells it is appropriate to identify the notion of self –

> "…it is legitimate to invoke the notion of "self" because the dynamics of the system is characterized by an invariant topological pattern that is recursively produced by the system and that defines an outside to which the system is actively and normatively related" (Thompson, 2007, 260).

And Thompson argues that this form of organization already implies a basic form of sense-making, cognition or incipient mind. He defines cognition as "behavior or conduct in relation to meaning or norms that the system itself enacts or brings forth on the basis of its autonomy" (Thompson, 2007, 126). Insofar as autopoiesis involves "internal self-production sufficient for constructive and interactive processes in relation to the environment, then, autopoiesis does entail cognition" (Thompson, 2007, 127). On this view then, life, even at the basic level of a cell, is not just physical in the standard materialist sense of purely external structure and function. Rather, "life realizes a kind of interiority, the interiority of selfhood and sense-making" (Thompson, 2007, 238). On this analysis then the kind of independence that we experience as a primary fact about the way we find ourselves living is present in all life from the first cells. I will return to this in chapter 6, section 5.

4.3 ORGANISMS

Maturana and Varela (1987, 87-89) make a distinction between living cells as first-order autopoietic systems and second-order autopoietic systems (multicellulars) where the systems include individual cells as structural components. Examples of multicellulars they list include organisms, colonies, and societies. This accords with Virchow's view (1978, [1858]) "What is an organism? A society of living cells, a tiny well-ordered state." This suggests an atomistic ontology in the biological sphere – but could any second-order autopoietic systems also be first-order autopoietic systems?

Thompson (2007, 106) argues that an autopoietic system is a specific kind of autonomous system – one having an organizational closure of production processes in the molecular domain. But there could also be other kinds of autonomous systems with different kinds of closure or boundary. The crucial aspect for any autonomous system is that it produces and regulates its own internal topology and functional boundary – the particular physical structure that realizes this boundary is not important.

A central question of biology is how distinct and heterogeneous components interact and constitute a cohesive whole, functioning collectively as a regulated unit that persists through time. The name for such a cohesive whole is an organism. Biologists have made differing attempts to specify the unity of functioning that characterizes living organisms – " the 'glue' that brings together the heterogeneous elements into a cohesive whole" (Pradeu, 2016, 804). C.M. Child (1915) introduced the concept of 'metabolism', which refers to all the chemical reactions involved in maintaining the living state of cells and organisms. Walter B Cannon (1926) introduced the idea of 'homeostasis' – the coordinated physiological reactions which actively maintain most of the steady states in the body, despite external changes. Claude Bernard (1974) identified 'nutrition' as the most crucial and characteristic phenomenon of life. By 'nutrition' he meant the process of destruction and construction that insures the persistence of a living being.

4.4 THE IMMUNE SYSTEM

While each of these different approaches is to some extent helpful, Thomas Pradeu (2012) argues that it is the immune system that is pivotal to understanding physiological individuality. He gives three arguments for this claim. Firstly, the immune system constitutes a discrimination mechanism, accepting some entities in the organism and rejecting others, thus participating decisively in the delineation of the organism's boundaries. Secondly, unlike many bodily 'systems (e.g. digestive or respiratory), the immune system is truly systemic (it exerts its influence everywhere in the organism). And thirdly, all living things have an immune system, including prokaryotes, plants, invertebrates and vertebrates, so an immunity-based account of biological individuality applies to the whole living world.

By analysing how immunologists have suggested the immune system is involved in the definition of biological identity, Pradeu identifies three senses or aspects of individuality. Firstly, uniqueness: what makes each biological individual unique? Secondly, delineation: what are the boundaries of a biological individual? Thirdly, persistence: what insures that something remains the 'same' biological individual despite constant change? Pradeu's three aspects of individuality correspond in broad outline to my three modes of existence in chapter 1 – independence, intersubjectivity, and interdependence. It's our individual uniqueness that makes us independent; each independent individual must delineate itself from others and the world; each individual has to maintain itself despite constant change due to our interdependence with the world.

The development of immunology as a scientific field was driven by the advent of transplantation medicine. This is how Burnet and Ferrer (1949) argued -

> "It is an obvious physiological necessity and a fact fully established by experiment that the body's own cells should not provoke antibody formation. […] An animal's own red cells are non-antigenic. This is not due to any intrinsic absence of antigenic components; the same cells injected into a different species or even into another unrelated animal

> of the same species may give rise to active antigen production. The failure of anti-body projection against autologous cells demands the postulation of an active ability […] to recognize 'self' pattern from 'not-self' pattern in organic material taken into their substance" (Burnet and Fenner, 1949, 85-86).

Burnet developed this idea as the self-nonself theory of immunity - " … the idea that the primary function of the immune system of an organism is the discrimination between what belongs to the organism – i.e. is of its self – and what does not, and that immunology is therefore the study of this discrimination" (ibid. p.1).

This theory grounds the question of the boundaries of a living thing in a question of origin: some things that come from the outside trigger a rejection response, while things that comes from the inside do not trigger an immune response. The theory seems to provide a plausible mechanism to integrate - the body's reaction to infectious agents, allergies, the rejection of transplanted organs and tissues, and, autoimmune diseases. On this view, immunology is understood as the study of the defence of living things against pathogens.

However, Pradeu (2016, 805) points out that research in immunology since the 1980's strongly suggests that this self-nonself framework is inadequate. It was found that self-constituents can trigger immune responses in healthy conditions, and many nonself constituents, including many microbes, are not rejected by the immune system. This lead Pradeu to conclude that immunological discrimination is not just based on origin.

Pradeu's alternative is what he originally called 'the continuity theory' (Pradeu, 2012), but what he now more accurately calls 'the discontinuity theory of immunity' (Pradeu and Vivier, 2016). According to this theory the basis for an immune reaction is "a strong modification of the antigenic patterns (ligands) with which the organism's immune receptors interact, which is to say a sudden appearance of antigenic patterns in the organism that differ strongly from those with which the immune system is continuously interacting" (Pradeu, 2012, 175). This suggests that the immune system works by detecting the rate of change in a stimulus rather

than the stimulus itself only. Therefore, according to the (dis)continuity theory – "what triggers an immune response of rejection is not genetic foreignness, but the expression of strongly unusual molecular patterns" (Pradeu, 2013, 10).

An advantage of this theory is that it is more specific. Immunogenicity is mapped to specific entities, such as receptors and ligands; and characteristics, such as unfamiliarity and suddenness of appearance (rather than conceptually vague terms such as self or danger which cannot be directly associated with concrete molecular or cellular markers). Also the discontinuity theory is able to explain a larger set of immunological reactions in a more consistent manner than the self-nonself theory, which often resorts to claiming exceptions e.g. autoimmune interactions, the reaction to tumors, the phagocytosis of dead cells of the body, the tolerance of a foetus by the mother, and, the survival of commensal bacteria in the gut.

Pradeu's theory has interesting consequences for the concepts of biological individuals or organisms. Whereas the older theory presupposed a self-nonself distinction based on origin, the discontinuity theory suggests that what counts as an individual is a consequence, (rather than a presupposition), of the immunological system. According to Pradeu, a multicellular organism is made of heterogeneous constituents, some of them present from conception or birth, while others appear in the organism throughout its life -

> "… every multicellular organism is *heterogeneous,* that is, made of genetically different entities. In particular, every multicellular organism (plant, invertebrate, vertebrate) hosts hundreds of billions of commensal and symbiotic bacteria (for example, it is estimated that the gut of a mammal hosts 10 to the power of 12 microorganisms). Due to the massive presence of symbionts, every multicellular organism is a *chimera*" (Pradeu, 2013, 10).

Pradeu argues that some symbiotic bacteria are not simply present in the organism but rather they are key functional *constituents* that play indispensible roles notably in digestion, immunity and development. They are massively present in all the interfaces of the organism and often play a

decisive evolutionary role, because they make a difference to the host's fitness and can be transmitted to the next generation.

Because organisms are so heterogeneous, they need a continuous mechanism of achieving unity or cohesiveness -

> "… a multicellular organism is characterized by the continuous construction of a cohesiveness through the permanent control of present constituents by the immune system. Immunity is key to insure the unity and cohesiveness of the organism. In my view, traditional evolutionary individuation needs to be complemented by an immunity-based individuation, because of the crucial role of the immune system in insuring the unity of different biological constituents. I suggest to call biological individuals that are cohesively organized through an immune system, 'organisms'" (Pradeu, 2013, 10-11).

He defines an organism as " a functionally integrated whole composed of heterogeneous components that are locally interconnected by strong biochemical interactions and controlled by constant systemic immune interactions of a constant average intensity" (Pradeu, 2012, 244).

Pradeu's theory shows how the immune system in organisms, such as ourselves, is the mechanism by which they maintain themselves as individuals in an ever changing environment. However if, as Pradeu argues, the immune system is primarily driven by discontinuity, it may need to be supplemented to explain the kind of self-producing individuals that we are.

4.5 AUTONOMOUS WHOLES

Pradeu's account of the immune system gives an account of the mechanism of integration of organisms. But in what sense is what is composed a whole over and above its components?

According to Peter Van Inwagen (1990. Sect. 12) organisms are the only real wholes. He claims that most ordinary material objects do not exist.

What we refer to as a chair is "a collection of simples arranged chairwise". The simples in question do not *compose* anything; there *is no* whole of which they are parts. Quite simply, there are no chairs. The only exception to this ban on composite objects is organisms. There really are organisms; the simples of which they are made really are parts that compose a whole. So the inventory of the physical universe, at its most general level, is as follows: there are simples, and there are organisms composed of simples. There is nothing else. If this is to be viable metaphysical account of organisms, we need some explanation of how, and in what sense, in the case of organisms, simples compose a whole.

Ruiz-Mirazo and Moreno (2004, 235) claim to give –

> "a bottom-up account of the origins of 'self-production' (or self-construction, as we propose to call it), pointing out which are the minimal material and energetic requirements for the constitution of basic autonomous systems. This account is, indeed, committed to the project of developing a general theory of biology, but well grounded in the universal laws of physics and chemistry."

They distinguish between-

- "spontaneous" dissipative structures, where the flow of energy and/or matter that keeps the system away from equilibrium is not controlled by the organization of the system (the key boundary conditions are externally established, either by the scientist in the lab or by some natural phenomenon that is not causally dependent on the self-organizing one).
- real autonomous systems, where the constraints that actually guide energy/matter flows from the environment through the constitutive processes of the system are endogenously created and maintained. To achieve this, the systems have to be chemical rather than purely physical.

They define basic autonomy as the capacity of a system to *manage* the flow of matter and energy through it so that it can, at the same time regulate, modify, and control (i) internal self-constructive processes and (ii) processes of exchange with the environment. They claim that this basic autonomy is evident at the level of the cell.

At higher levels, cohesion of the system has to be achieved by functional coordination (Moreno, 2016). When a new system is formed additional system-related functions are required. This requires a hierarchical, regulated organization of the constitutive and interactive functions under the control of a dedicated subsystem, which operates at a higher level. This regulatory center constitutes an integrated "self", and generates a form of agency capable ensuring the maintenance and development of the system. A condition for reaching full individuality is symmetry between the interactive and constitutive components.

At the constitutive level, such a system requires genetic homogeneity, functional plasticity, metabolic capacity and an immune system. At the interactive level, it requires a nervous system. Together the immune system and the nervous system form a global regulatory system – the self. When the right sort of function integration is achieved, a cohesive individual emerges which can be subject to evolutionary selection.

Moreno (2016) points out that functional integration is a matter of degree. You have to examine the parts and the whole to see where the stronger integration occurs. In practice this may be difficult to determine in composite systems such as termite colonies. The identification of a self is a matter of identifying the source of agency - the self involves top down control.

4.6 HYLOMORPHISM

Ruiz-Mirazo and Moreno's (2004) claim, then, is that an account of an emergent self can be given "grounded in the universal laws of physics and chemistry" – by what they claim is "a proper naturalization of the concept of autonomy". In Seager's terms they claim to have given a sufficient account of the kind of self we are in terms of conservative rather than radical emergence.

However Denis Walsh (2016) claims to identify a problem with this, at least if it is based on the currently widely accepted metaphysical position of

foundational materialism. A central claim of this metaphysical system is the Causal Inheritance Principle, which states that the causal properties of a complex entity are fixed by the properties of its parts. The problem with this, as Kim (2006, 548) pointed out, is that it makes the phenomenon of radical emergence incoherent, because radical emergence requires genuine downward causation. But if the Causal Inheritance Principle holds, then as John Searle (1984, 94) puts it –

> "Since all of the surface features of the world are entirely caused by and realized in systems of micro-elements, the behavior of micro-elements is sufficient to determine everything that happens. Such a 'bottom up' picture of the world allows for top down causation (our minds, for example, can affect our bodies). But top-down causation only works because the top down level is already caused by and realized in the bottom levels."

Top-down causation that only works because the top down level is already caused by and realized in the bottom levels is not genuine downward causation, and genuine downward causation is necessary to account for the kind of self we know ourself to be.

Denis Walsh (2016) argues however that the causal inheritance principle and downward causation are incompatible only because of a prior commitment to the metaphysical assumptions of foundational materialism. Foundational materialism is the conception of what it is to be an entity and the principles of change that were developed at the time of the scientific revolution. The main tenets of this position are, that a primary substance is constituted exclusively of its matter, and that the nature of complex individuals is determined exclusively by the matter that composes them. Walsh argues that this metaphysical position has become a genuine impediment to the proper understanding of the nature and functions of biological systems, in which the function of the system confers certain causal properties upon its parts.

Walsh therefore recommends that we should reconsider the earlier Aristotelian metaphysics of hylomorphism. According to hylomorphism, a primary substance is a complex of matter and form. Both must be specified –

but form is explanatorily primary – form is a restriction on matter. Maurice Merleau-Ponty explains the concept of form as a whole that cannot be dislocated from its components but cannot be reduced to them either. As he puts it - "there is form whenever the properties of a system are modified by every change brought about in a single one of its parts and, on the contrary are conserved when they all change while maintaining the same relationship among themselves" (Merleau-Ponty, 1963, 47).

As William Jaworski (2016) argues, structure (or organization, form, arrangement, order, or configuration) is a basic ontological and explanatory principle; it concerns what things there are, and, it concerns what things can do.

> "Some individuals, paradigmatically living things, consist of materials that are structured or organized in various ways. You and I are not mere quantities of physical materials; we are quantities of physical materials with a certain organization or structure. That structure is responsible for us being and persisting as humans, and it is responsible for us having the particular developmental, metabolic, reproductive, perceptive and cognitive capacities we have" (Jaworski, 2016, 1).

Jaworski thus argues that composite individuals are *emergent,* in that they have distinctive powers different from the powers of unstructured materials. He argues that physics, insofar as it views the world as a vast sea of matter and energy, cannot account for emergent individuals. "[I]t does not enable us to describe the overall pattern that makes this collection of fundamental physical interactions a human, a dog, or an oak tree". "It won't tell us … what sets the whole apart as a unified individual distinct from the inanimate materials that surround it," and, "the materials by themselves won't tell us why the whole possesses powers that the materials by themselves do not possess, such as the powers to perceive, or think, or feel" (Jaworski, 2016, 25).

Therefore he argues that there must be some further principle that accounts for the unity and also the persistence of the whole that explains why numerically one and the same whole can exist at different times despite

the constant flux of materials that are caught up in it, and that explains the distinctive powers of living things. "That principle, say hylomorphists, is structure, the persistent dynamic structures we find in living things" (Jarorski, 2016, 25).

On this view, the matter/form distinction is relational rather than foundational; the part/whole relation is reciprocal rather than compositional; and the properties of matter are contextual rather than intrinsic. Walsh argues that hylomorphism provides an appropriate model for the individuation and explanation of biological systems.

Hylomorphism, at least as presented by Jaworski, entails emergentism in what Seager calls the radical sense. As I argued in chapter 3, radical emergence is inexplicable in a purely physical account of the world. It seems we must accept that there are some emergent features that we must include as brute facts. Similarly, hylomorphism argues that physics cannot account for emergent individuals, but they can be explained in terms of 'structure'; but 'structure' is just "the persistent dynamic structures we find in living things". So we can find instances of structure but there is no physical explanation of why this structure occurs here. But, as I argue in the next section, the fact that *we* can find them is significant.

4.7 KNOWING LIFE BY BEING ALIVE

As I have argued above autopoiesis at the level of the cell, and autonomy at the level of the organism, gives rise to inwardness and outwardness – the self-production of an inside that also specifies an outside to which it is normatively related. This is the structure of living things. Hans Jonas (1966) suggests that this structure of inwardness or interiority is disclosable to us because we ourselves are living beings who experience our own bodily selfhood at firsthand. He argues that from the disembodied, analytical standpoint, the organism would be resolved into a collection of fleeting, objective physicochemical events, "and all features of a self-related autonomous unity would, in the end, appear as purely … fictitious" (Jonas,

1996, 78). However, as bodily beings ourselves, we experience inwardness and purposiveness in the way we find ourselves living in the world – "we happen to have inside knowledge" (Jonas, 1996, 79). Therefore Jonas argues "*we* are able to say what no disembodied onlooker would have cause for saying: that the mathematical God in his homogeneous analytical view misses the decisive point – the point of life itself: its being self-centered individuality, being for itself and in contraposition to all the rest of the world, with an essential boundary dividing 'inside' and 'outside' " (Jonas, 1996, 79). Jonas summarizes his thinking with the proposition – "life can only be known by life" (Jonas, 1996, 91).

Evan Thompson considers Jonas's proposition as a quintessentially phenomenological one – "before being scientists we are first and foremost living beings, and we thus possess within ourselves evidence of our purposiveness" (Thompson, 2007, 163). Purposiveness or teleology is a primary fact about ourselves that we know from the way we find ourselves living. Thompson also argues that the proposition is transcendental in a phenomenological sense – it is about the conditions of the possibility of knowing life, given that we do actually have biological knowledge. He argues that to account for certain observable phenomena we need the concepts of organism and autopoiesis. And, "these concepts and the biological accounts in which they figure are not derivable from some observer-independent, nonindexical, objective, physico-chemical description, as the physicalist myth of science would have us believe" (Thompson, 2007, 164). Rather, the source for the meaning of these concepts is the lived body, our original experience of our own bodily existence.

As Merleau-Ponty put it – "I cannot understand the function of the living body except by enacting it myself, and except in so far as I am a body which rises towards the world" (Merleau-Ponty, 1962, 75).

Thompson argues that this transcendental perspective overturns the uncritical standpoint of objectivist philosophy and science.
"Objectivism takes things for granted, without asking how they are disclosable to human experience and knowledge, or how they come to be disclosed with the meaning or significance they have" (Thompson, 2007,

164). Specifically, Thompson argues that objectivism in biology takes the organism for granted as a ready-made object out there in the world, without considering how the category 'organism' is constituted for us in scientific experience. Phenomenology, on the other hand, traces this category back to its cognitive source, which is the lived experience of our bodily being.

This argument seems to be particularly applicable to biology, but Merleau-Ponty argues that it applies to all science –

> "The whole universe of science is built upon the world as directly experienced, and if we want to subject science itself to rigorous scrutiny and arrive at a precise assessment of its meaning and scope, we must begin by reawakening the basic experience of the world of which science is the second-order expression" (Merleau-Ponty, 1962, viii).

Thompson (2007) argues that a critical and reflective science should embrace this phenomenological perspective because, by clarifying scientific experience in this way, science itself can be properly situated in relation to the rest of human life and thereby be secured on a sounder footing. This is the topic of the next chapter.

CHAPTER 5. SCIENTISM & THE METAPHYSICS OF BEING HERE

5.1 Modifying metaphysics
5.2 Scientism as metaphysics
5.3 Re-visioning the nature and role of science
5.4 Phenomenology and science
5.5 The 'problem' of consciousness
5.6 Dissolving the hard problem
5.7 The metaphysics of being here

5.1 MODIFYING METAPHYSICS

In Chapter 3 we saw that Seager (2012, 2014) identifies major problems with his first three options for integrating consciousness into the scientific picture of the world. 'Watchful waiting' may be like a person walking towards the horizon with the hope of reaching it: just as a horizon is not an object in the world that can be reached by walking towards it, consciousness may not be the kind of entity that can be fitted in to an objective picture of the world. The two more radical options of 'embracing emergence' and 'favouring fundamentality' add elements to the picture of the world for which there is no epistemic justification, and seem to be mirror images of each other – the strengths of each is seen as the weakness of the other. Without some account of consciousness at the fundamental level, radical emergence seems like a brute fact (or 'magic') that we must accept with 'natural piety'. On the other hand, any account in terms of fundamental consciousness entities ('mind-dust') faces a problem of explaining how these conscious entities combine (or in the case of cosmopsychism – subdivide) to form the conscious subjects we know by being one. I argue that these apparent problems justify setting aside these theories to consider an even more radical option.

In Chapter 4 I argued that many of the concepts that the science of biology uses are based on and presuppose the knowledge we have of ourselves as

living beings. The impossibility of giving a purely objective account of life suggests a re-ordering of priority self and the science of biology.

In this chapter, I will consider if this lesson from biology can be applied more widely to science in general. This is what Seager suggests is the fourth option for integrating consciousness and the scientific picture of the world. He calls it 'modifying metaphysics', because it involves a "radical re-visioning of the nature and role of science", such that the scientific picture of the world, as traditionally conceived, is no longer understood as the complete description of reality. On this view, ultimate reality is not to be sought in anything transcending experience; rather the world is taken to be as it is given in our ordinary experience. Science plays a particular role and adopts a specific third person method to investigate aspects of our world, but it is not our only way of understanding our world. In so far as it dismisses or ignores our first person experience, science is limited as a source of metaphysics.

5.2 SCIENTISM AS METAPHYSICS

A dominant assumption of much of analytic philosophy is scientific rationalism and its underlying monism. As Isaiah Berlin put it, the assumption is that "there is but one reality and one proper description of that reality, for which there is an appropriate language into which all knowledge statements can be translated, otherwise such statements are false or meaningless" (Berlin, 1979, 80-81). Science, and especially the science of physics, is now widely held, by scientists, philosophers, and our culture more generally, as the sole source of knowledge about the ultimate nature of reality. Particle physics is promoted because of " its unique role in discovery of the fundamental nature of the universe" (Banegas et al., 2007).

Following in this tradition, Ladyman and Ross (2007, 27) argue that one of the important things that we want from science is a relatively unified picture of the world. They claim that this is not asserted as a primitive norm,

but rather the justification for this is from within the history of science, in that scientists are reluctant to pose or to accept hypotheses that are stranded from the otherwise connected body of scientific beliefs (Kitcher, 1981). Also, as Thagard (1992) argues, an important source of justification for a hypothesis is its standing in reciprocal explanatory relationships – networked consilience relationships – with other hypotheses. Because evaluating consilience networks across the sciences cannot be achieved by science itself, Ladyman and Ross accept the need for some metaphysics. However, they say any metaphysics must be motivated exclusively by attempts to unify hypotheses and theories that are taken seriously by contemporary science. So, although metaphysics is an attempt to go beyond science to achieve an overall picture of reality, it must always be grounded in and not contradict the results of science. To ensure this they propose the following two verificationist principles –

- "No hypothesis that the approximately consensual current scientific picture declares to be beyond our capacity to investigate should be taken seriously."
- "Any metaphysical hypothesis that is to be taken seriously should have some identifiable bearing on the relationship between at least two relatively *specific* hypotheses that are either regarded as confirmed by institutionally *bona fide* current science or are regarded as motivated and in principle confirmable by such science" (Ladyman and Ross, 2007, 29).

Ladyman and Ross combine these principles with their *Primacy of Physics Constraint,* which I mentioned in section 3.4. This constraint articulates the sense in which evidence acceptable to naturalists confers epistemic priority on physics over other sciences. These considerations motivate Ladyman and Ross's *'Principle of Naturalistic Closure'*, which aims to be a naturalistic constraint on metaphysical hypotheses –

> "Any new metaphysical claim that is to be taken seriously … should be motivated by, and only by, the service it would perform, if true, in showing how two or more specific scientific hypotheses, at least one of which is drawn from fundamental physics, jointly explain more than

> the sum of what is explained by the two hypotheses taken separately…" (Ladyman and Ross, 2007, 37).

This principle, then, aims to reduce metaphysics merely to the task of unifying scientific hypotheses. However, Tuomas Tahko (2012, 37) argues that this ignores a crucial element, in that metaphysical inquiry is required in order to produce scientific hypotheses in the first place. He points out that the reasoning process, which leads to the formulation of a scientific hypothesis fulfils all the essential features of metaphysical inquiry. This suggests that paying attention to our own reasoning processes, which we know by using them, should be prior to any inquiry, whether scientific or metaphysical.

Ladyman and Ross's trust in science is based on a suspicion about the reliability of our individual experience to track reality. And this applies particularly to philosophers' individual experiences: as they put it –

> "… it seems to us to be just ridiculous when philosophers look up from their desks and tell us that while sitting there and concentrating they've discovered (usually all by themselves) facts about the nature of the world that compete with the fruits of ingenious experimentation conducted under competitive pressure and organized by complex institutional processes" (Ladyman and Ross, 2007, 57).

On the basis of evolution, we might expect, as Willard Quine (1969) hypothesized, that the evolutionary processes that designed people should have endowed us with cognition that reliably tracks truth, on the grounds that believing truth is in general more conducive to fitness than believing falsehood. Ladyman and Ross accept that this may well be true, but they point out that this does not imply that our everyday or habitual intuitions and cognition are likely to track truths across all domains of enquiry. They argue that human intelligence evolved mainly to enable us to navigate complex social coordination games. Also, people are probably relatively reliable barometers of the behavioural patterns of animals they get to spend time observing, at making navigational inferences in certain sorts of environments, and at anticipating aspects of the trajectories of medium-sized objects moving at medium speeds. However, they argue that,

"proficiency in inferring the large-scale and small-scale structure of our immediate environment, or any feature of parts of the universe distant from our ancestral stomping grounds, was of no relevance to our ancestors' reproductive fitness. Hence, there is no reason to imagine that our habitual intuitions and inferential responses are well designed for science or metaphysics" (Ladyman and Ross, 2007, 2).

Their contention, then, is that the domain of science and metaphysics is much deeper and broader than our everyday world. Our inherent intuitions therefore may not be reliable guides to the structure of reality as a whole. However, by representing the world and reasoning mathematically people can abstract away from their familiar environment, and thus achieve scientific knowledge. Since this knowledge can be incorporated into unified pictures, we also can have some justified metaphysics (although because our current science is incomplete and may be found to be ultimately false, our current metaphysics will also be incomplete and ultimately false).

> Philip Goff characterizes Ladyman and Ross's 'big picture' as "Methodological naturalism – The lesson we should draw from the success of natural science is that we should look to, and only to, the third-person scientific method (i.e. rigorous empirical investigation of what is publically observable) to tell us what reality is like" (Goff, 2017, 5).

Ladyman and Ross suggest then, that in order to understand the metaphysics of reality we need to pay less attention to our immediate lived world, which can lead us astray, and more attention to the very large, the very small and the distant, which are best investigated by objective science. However, looking at the actual results of science might suggest that this should not be the only strategy.

As the cosmologist and astrophysicist Martin Rees (2017) points out – "Scientific knowledge is actually surprisingly 'patchy' – and the deepest mysteries often lie close by." For example, by using science we can convincingly interpret measurements that reveal two black holes crashing together more than a billion light years from Earth, but we have made little progress in treating the common cold, despite great leaps forward in

epidemiology. Rees argues that this is not as paradoxical as it looks because "Astronomy is far simpler than the biological and human sciences. Black holes, although they seem exotic to us, are among the uncomplicated entities in nature. They can be described exactly by simple equations … black holes are simpler than forests".

Thinking in terms of complexity, then, might suggest a different focus for metaphysical enquiry. As Rees puts it – "Atoms and astronomical phenomena – the very small and the very large – can be quite basic. It's everything in between that gets tricky". The most complex things are living entities and the most complex living entities are humans. Rees points out that "an animal has internal structure on every scale, from the proteins in single cells right up to limbs and major organs. It doesn't exist if it is chopped up, the way a salt crystal continues to exist when it is sliced and diced. It dies." In trying to get an overall picture of reality, the close and complex may be more important than the micro, cosmic and distant.

But, complexity in itself might not be an issue beyond the practical difficulty of calculation. Physicist Steven Weinberg (1992) argues that all explanatory arrows point downwards. If you keep asking 'why' questions, you end up on the level of fundamental particles or fields. On this view, everything, however complex, is a solution to Schrodinger's equation – the basic equation that governs how a system behaves, according to quantum theory.

However, this type of reductionist explanation isn't always the best or most useful. Although everything may be made of atoms or sub-atomic fields, most scientists are interested in macroscopic systems that manifest properties that are best understood by concepts and "patterns" appropriate to the level of the system. Thus, the flow of water in rivers is better understood in terms of viscosity and turbulence, rather than by the interaction of individual atoms, and even temperature (i.e. mean molecular kinetic energy) is inapplicable at the scale of individual atoms. . Martin Rees (2017) points out that only about 1% of scientists are particle physicists or cosmologists – the other 99% work on 'higher' level macroscopic systems. These macroscopic systems seem to be invisible from the particle or field

level; but, being one ourself, they seem very useful and significant to us. As I discussed in the previous two chapters, the reality of these macroscopic features has been understood philosophically in terms of 'emergent' properties or entities; or by hylomorphists in terms of the fundamentality of structure.

Ladyman and Ross explain the reality of macroscopic features in terms of an ontology that is 'scale relative'. According to this idea "claims about what (really, mind-independently) exists should be relativized to (real, mind independent) scales at which nature is measureable" (Ladyman and Ross, 2007, 200). They use and develop Dan Dennett's (1991) idea of real patterns (rather than individuals) as the entities that exist at different scales, and they describe their analysis of this pattern-reality as 'rainforest realism', to express their idea that ontology is extensive and diverse, in contrast to Quine's (1953) 'desert landscapes'.

In spite of, or maybe, because of this diverse ontology, Ladyman and Ross seem strangely noncommittal or uninterested in the ontological status of persons or selves, who are presumably complex real patterns that can identify and use other real patterns. They concede that "our concept of selves may turn out to track a real pattern, but it may also attribute properties to selves that the real patterns lack", and they say that "If cognitive science concludes that mental concepts do not track any real patterns then the theory of mind will have to go" (Ladyman and Ross, 2007, 254). Science determines ontology but it seems, on their view, the scientist has no special status.

However, the significance of complexity suggests that metaphysics should be particularly interested in sciences that investigate living and particularly human 'patterns'. Moreover, prior to any such scientific knowledge, our first person knowledge of ourselves is the basis of all knowledge and therefore is significant for any metaphysical knowledge of reality.

Michel Bitbol and Claire Petitmengin (2013) explain how first-person reports do have a crucial epistemic role. They acknowledge that first-person reports by single individuals may indeed be flawed when they are taken at face value, as exhaustive descriptions of, and objective knowledge *about,* the

cognitive processes taking place in these individuals, and objective knowledge *about* the world. But although first-person reports may fail to be self-sufficient pieces of knowledge (beyond acquaintance), they remain the unique and inescapable basis of any further empirical knowledge of ourselves, and our environment.

According to this view, first-person access is –

(a) the testimony of our being-in-the-world, and

(b) the source of every claim of the availability of a surrounding world.

(c) Indispensible to ascribe functional meaning to most neurophysiological patterns. As Seager (2012, 166) points out while patterns are there to be seen they have no function if they are not seen, and neurophysiological patterns, in particular, can only be 'seen' and therefore understood from the first-person perspective.

(d) The source of all 'objective' knowledge – what we take to be the 'objective experimental data' of natural sciences just is the intersubjective convergence of (a large number of) *first-person reports* of a certain type. As Bitbol and Petitmengin (2013) point out – "What gives objective data or measurements their reliability is nothing else than the coarseness of the categories in which first-person reports are constrained to fall, assisted by instrumental amplification of coarseness".

If this is correct then reliability does not require the complete *elimination* of first-person reports. Rather first-person reports remain the *de facto* starting point and ultimate warrant of the whole system of knowledge.

5.3 RE-VISIONING THE NATURE AND ROLE OF SCIENCE

Modifying metaphysics is primarily about the nature of reality. It involves giving up on ultra-strong scientific realism (scientism) – the view that science (especially physics) is the sole source of knowledge about the ultimate nature of reality. Frank, Gleiser and Thompson (2019) characterize this belief that physical reality has absolute primacy in human knowledge – as *scientific materialism.* Scientific materialism combines scientific

objectivism (the claim that science tells us about the real, mind-independent world) and *physicalism* (the claim that science tells us that physical reality is all there is). Frank, Gleiser and Thompson identify two intractable problems with this framework. Objectivism is problematic because we never encounter physical reality outside of our observation of it – "Elementary particles, time, genes and the brain are manifest to us only through our measurements, models and manipulations. Their presence is always based on scientific investigations, which occur only in the field of our experience."

Physicalism is problematic because, as we saw in chapter 3, Hempel argued that it is either false (because current science doesn't include an account of experience, and we know experience exists) or empty (because we do not know what future physics will say particularly about consciousness). In any case, Frank, Gleiser and Thompson (2019) point out that objectivism and physicalism are philosophical ideas, not scientific ones – even if some scientists espouse them. Objectivism and physicalism don't necessarily follow from what science tells us about the physical world, or from the scientific method itself. They are at best meta-scientific. By not realizing that these perspectives are a philosophical bias or interpretation, not directly supported by science, scientific materialists ignore the ways that immediate experience and the world can never be separated.

As Arthur Schopenhauer (1818/9, 26-27) argued materialism is "the attempt to explain what is directly given to us from what is given indirectly." He points out that materialism tries to find the first and simplest state of matter, and then to develop all the other states from it. So we start from "mere mechanism" and try and develop chemistry, the vegetable and the animal kingdoms, and ultimately sentience from this. But Schopenhauer argues that sentience was already presupposed as the indispensable condition at the very first starting-point, at mere matter: without sentience there can be no attempt to find the simplest state of matter. As Giulio Tononi (2015) puts it "the existence of one's consciousness and its other essential properties is certain, whereas the existence and properties of the physical

world are conjectures, though very good ones, made from within our own consciousness".

In the alternative re-visioning or re-framing, then, the world is taken to be as it is given in our ordinary experience. Ultimate reality is not to be sought in anything over and above experience. On this view then, science can help us to understand the structure and processes in this world, but this help is in the form of a large set of diverse models, which are derived from, and therefore secondary to, observation and experimentation. Frank, Gleiser and Thompson (2019) identify five steps in the scientific method –

- First, we set aside aspects of human experience on which we can't always agree, such as how things look or taste or feel.
- Second, using mathematics and logic, we construct abstract, formal models that we treat as stable objects of public consensus.
- Third, we intervene in the course of events by isolating and controlling things that we can perceive and manipulate.
- Fourth, we use these abstract models and concrete interventions to calculate future events.

Fifth, we check these predicted events against our perceptions.

Frank, Gleiser and Thompson point out that experience is involved in every step of this scientific process, and that the scientific models are abstractions that must be pulled out from observations –

> "They are idealizations, not actual things in the world. Galileo's model of a frictionless plane, for example; the Bohr model of the atom with a small, dense nucleus with electrons circling around it in quantised orbits like planets around a sun; evolutionary models of isolated populations – all of these exist in the scientist's mind not in nature. They are abstract mental representations, not mind-independent entities" (Frank, Gleiser and Thompson, 2019).

Thus, scientific explanations are model based, and it is presumed that the models are but imperfect replicas of small portions of reality, which are tailored to the specific job of prediction they are designed to aid. Models that quantify their subject matter are preferred because that they are easier to use and result in greater predictive power. However, this doesn't mean that

our scientific knowledge is arbitrary, or just a projection of our own minds. Some models and explanations work much better than others and we can test this. But testing this does not give us access to nature as it is in itself, outside of our ways of seeing and acting on things. Therefore, these models do not point to a comprehensive view of the world or an accurate picture of ultimate reality. Thus, Frank, Gleiser and Thompson (2019) argue that the belief that scientific models correspond to how things truly are doesn't follow from the scientific method. Rather, they argue that "it comes from an ancient impulse – one often found in monotheistic religions – to know the world as it is in itself, as God does. The contention that science reveals a perfectly objective 'reality' is more theological than scientific."

On this view then, scientific models never take us outside experience, because they require specific kinds of perceptions performed by highly trained observers.

Scientific objectivity is not about accessing mind independent reality; rather 'objective' in this context, is better understood as meaning something that's true to the observations agreed upon by a community of investigators using certain tools.

In such a reoriented vision of science, consciousness takes its place as part of reality many of whose features can be scientifically explained. We can investigate the neural correlates of consciousness, the physical conditions necessary for consciousness to appear, the evolution of consciousness, etc. However, as Seager points out, on this view –

> "… the fact – if it is a fact – that the phenomenal or essentially subjective aspect of consciousness cannot be explained scientifically, nor integrated into some putatively total and comprehensive picture of reality based entirely on fundamental physics, is not a deep worry. Not everything in the world is amenable to the scientific modeling project" (Seager, 2014, 22).

This implies that any purely *scientific* picture of the world is never going to give us a complete picture of reality.

This view of science is more widespread than is sometimes acknowledged. It is broadly consistent with the position of Bas van Fraassen's constructive

empiricism (1980, 2002), which argues that empirical adequacy rather than realism is the aim of scientific theorizing. According to van Frasaasen, the belief that the theory fits the observable phenomena is the only belief involved in accepting a scientific theory; explanatory power is not grounds for believing that all theoretical terms refer. Nancy Cartwrights's (1983, 1999) views on the laws of nature (that they lie), and John Dupré's account of the disunity of the world, and therefore science (1993), also seem consistent with this view. A similar conception of science is found in Steven Horst's (2014) theory of 'cognitive pluralism'. According to Horst science does not passively reflect the world like Richard Rorty's (1979) mirror of nature, but rather science actively models specific features of it, one (or a few) at a time.

As Horst explains it, modeling features of the world involves both abstractions and idealizations. Abstraction involves screening out some features of the world to bring others into focus, like viewing an object through a particular type of lens. Idealization is a special type of abstraction which can involve 'bracketing' – screening out features that matter in vivo, or 'distorting' – for example treating objects as point-masses. As Nancy Cartwright (1983) argued 'the laws of physics lie' in that they never exactly describe the behavior of objects in the real world: rather they describe the behavior of highly idealized objects in models. Thus, for example, real world objects never behave as gravitational theory, taken alone, would predict. And Horst says that scientific models are not only idealized, but also, optimized for particular pragmatic goals, and subject to constraints of our human cognitive architecture. Different models arise because any model involves a choice of some particular representational system, involving possibly some guiding metaphor and a choice of a particular mathematical tool. Thus, wave and particle models capture different aspects of the phenomena of light by employing different representational systems. And the choice of which representational system that is chosen in any particular case will depend on what particular aspect the scientist is interested in.

But as Horst points out, these scientific models "can, nonetheless, be

models of the same reality … much as one can take pictures of a single object from different angles and through different camera lenses. However, pasting together such snapshots does not produce, and could never produce, a master picture of how the object looks from all angles at once. Pictures do not work like that, and neither, quite plausibly, do theories or models" (Horst, 2014, 214).

On this view then, our scientific 'snapshots' are genuine, but have limited usefulness in producing a metaphysical 'master picture'.

5.4 PHENOMENOLOGY AND SCIENCE

Seager's option of modifying metaphysics is also very similar to the approach that has been advocated by phenomenologists going back to Edmund Husserl.

In "The Crisis of European Sciences and Transcendental Phenomenology", written in the 1930s, Husserl criticized the confining of our world-view or philosophy to science, and the positivistic reduction of the idea of science to mere factual science. He claims that the total world-view of modern man has "let itself be determined by the positive sciences" and this has led to "an indifferent turning away from the questions which are decisive for genuine humanity. Merely fact-minded sciences make merely fact-minded people." Science "excludes in principle precisely the questions which man . . . finds the most burning: questions of the meaning or meaninglessness of the whole of human existence" (Husserl, 1936/1970, 6).

Husserl points out that the idea of philosophy developed in the Renaissance from ancient sources, was much broader. It envisaged philosophy as "one all-encompassing science, the science of the totality of what is" (Husserl, 1936/1970, 8).

But the great success of the positive sciences led to skepticism about the possibility of metaphysics and the collapse of the belief in a universal philosophy. Husserl argues that this has led to a collapse of the belief in

reason. He says that it is "reason which ultimately gives meaning to everything that is thought to be, all things, values, and ends" (Husserl, 1936/1970, 12). He says that "if man loses this faith, it means nothing less than the loss of faith 'in himself', in his own true being" (Husserl, 1936/1970, 13).

Husserl calls for the reestablishment of philosophy with a new universal task. He says that philosophy and science should be "the historical movement through which universal reason, 'inborn' in humanity as such, is revealed" (Husserl, 1936/1970, 16). A universal philosophy would encompass different areas of investigation for which differing methods and concepts would be appropriate.

Merleau-Ponty takes up this question in the Preface to *Phenomenology of Perception*. He rejects the scientific view of humans as an adequate description of himself –

> "I am not the outcome or the meeting point of numerous causal agencies which determine my bodily or psychological make-up. I cannot conceive of myself as nothing but a bit of the world, a mere object of biological, psychological or sociological investigation" (Merleau-Ponty, 1962, viii).

Merleau-Ponty's contention is that science is a second order rationale or explanation, which is based on a more basic subjective experience of the world -

> "All my knowledge of the world, even my scientific knowledge, is gained from my own particular point of view, or from some experience of the world without which the symbols of science would be meaningless" (Merleau-Ponty, 1962, viii).

In order to be able to assess the proper role of science, we must begin by reawakening the basic experience of the world. Husserl saw phenomenology as a return to the 'things themselves'. Thus Merleau-Ponty says –

> "To return to the things themselves is to return to the world which precedes knowledge, of which knowledge always speaks, and in relation to which every scientific schematization is an abstract and derivative sign-language, as is geography in relation to the countryside in which we

have learnt beforehand what a forest, a prairie or a river is" (Merleau-Ponty, 1962, ix).

Merleau-Ponty thought that phenomenology had a role in showing science that its theoretical constructs are built on the foundation of our 'operative intentionality'. He claimed that scientific and logical 'objects' are abstractions of originally lived 'phenomena', rather than the other way around as is often thought in the 'scientistic' or 'scientific naturalist' approach.

Michel Bitbol (2008) argues that an analysis of ordinary and scientific knowledge shows us that objective domains of knowledge are elaborated in two steps, with conscious experience as an implicit departure point –

- Firstly, one progressively pushes aside any feature of experience on which conscious subjects cannot agree, such as individual tastes, community values, or the emotional tinge, which is associated, by individuals and communities with particular situations.
- Secondly, one only retains a sort of structural residue of conscious experience that can be the object of a consensus, and of a collectively efficient use as a predictive tool. At its most abstract, this structural residue is mathematical; but it can also consist of general propositions stating various types of relation between entities and predicates, such as "brains are bioelectrical organs made of neurons and glial cells.

However, as Bitbol argues, consciousness is and remains methodologically primary – "Any objective descriptions arise, in history and on a day-to-day basis as well, as an invariant structural focus for subjects endowed with conscious experience" (Bitbol, 2002).

But Bitbol (2008) claims that the very success of this procedure of extracting invariants yields a sort of amnesia - "The creators of objective knowledge become so impressed by its efficacy that they tend to forget or to minimize that conscious experience is its starting point and its permanent requirement … They finally turn the whole procedure upside down, by claiming that experience can be explained by one of its structural residues" (Bitbol, 2008, 56).

5.5 THE 'PROBLEM' OF CONSCIOUSNESS

Perhaps the mostly widely accepted reaction to the problem of fitting consciousness into the scientific picture of the world is to suggest that we wait and see what future science discovers. The history of science suggests that 'outlying facts' or 'exceptions' are eventually included within accepted theories as science becomes more and more comprehensive. As we saw in chapter 3, this conservative or "watchful waiting" (Seager, 2014,17) strategy assumes that whatever consciousness is, it will turn out to be a conservatively emergent phenomenon that will eventually be fitted into the scientific picture of the world in the standard way.

Many aspects of consciousness are already being investigated by science: for example the neural correlates of consciousness, the physical conditions necessary for consciousness to appear, or the evolution of consciousness. However, at least some aspects of consciousness as we live it seem to be prior to any possible description (scientific or otherwise) of anything at all. As Seager observes -

> "The problem of consciousness does not revolve around descriptions of consciousness but around the simple fact that conscious beings are presented with the world, and themselves, in a special way quite different from the causal and information laden reactions of more ordinary physical objects" (Seager, 2014, 15).

This aspect of consciousness - the simple fact of the way we are presented with the world and ourselves – is what Arthur Deikman (1996) terms awareness. "Awareness is the ground in which the mind's contents manifest themselves; they appear in it and disappear once again". It is of a different nature than the contents of the mind; it is prior to contents, more fundamental. As Husserl (1993) pointed out that there is a major difference of 'essence', between natural bodily objects and 'lived experience'; natural objects are always incompletely given, whereas lived experience is immediately and completely given – its existence is indubitable whenever it is present. As John Searle (1997) observes, "where consciousness is concerned, the existence of the appearance [just] is the reality".

Consciousness is the very fact that there is appearance; appearance is the reality of consciousness. As I mentioned in the introduction, this is what Michel Bitbol terms the radical self-referentiality of consciousness – "a question about consciousness is seen only if attention shifts from the meaning of the question to the present conscious experience as the background of this very act of attention" (Bitbol and Luisi, 2011, 210). This second focus is not an 'object'; rather, it is the condition for anything to be taken as an object.

Bitbol illustrates this concept of consciousness by considering how to answer the question "*where* does consciousness come from?" 'Where' questions are normally answered by pointing towards an object or a process that we can describe, think about, or at least imagine. But if, as Bitbol argues, consciousness is the background condition of any act of attention, then he argues that we are bound to reply that "consciousness comes from nowhere else than … *here*". But 'here' does not refer to any particular place, object or process. Rather 'here' is an indexical term, like 'I', 'now', and 'this'. Any such indexical term fully commits the person who utters it. In particular, 'here' is a verbal operator that brings each one of us back to one's own situation. Thus Bitbol argues that saying "consciousness comes from *here*" means that consciousness has no other obvious origin than the present situation (Bitbol, 2011, 210 -211).

This concept of consciousness is problematic for language and therefore for science, because science inevitably uses language and symbols. Rhawn Joseph (1982) points out two elementary features of language: language *means* and *discriminates.* Meaning displaces attention from the sound of a word to what it refers – it pushes us outwards towards something that is not close at hand. Therefore, when we use a word for 'consciousness', we are inevitably led astray, because conscious experience is not some*thing* over *there* to be *meant* in any way. Rather, as Bitbol argues "consciousness is plainly *here*; this 'here' that submerges us: this 'here' that is presupposed by any location in space". Similarly, Bitbol argues that the discriminative aspect of language cannot allow for the radically self-referential concept of consciousness. He argues that "*at this precise moment* [my consciousness]

contains in it all the memories of the past and all the projects for the future, [so that] there is nothing that can be contrasted with it" (Bitbol, 2011, 212).

Furthermore, as Edmond Wright (2008) points out, science was born from the decision to *objectify,* that is to select the elements of experience that are invariant across persons and situations. Its aim is to formulate *universal* truths, that is, truths that can be accepted by anyone irrespective of one's situation. Science is therefore concerned with the necessary connections between phenomena, but has little to say about the mere fact that there are phenomena or the qualitative content of these phenomena.

5.6 DISSOLVING THE HARD PROBLEM

According to this line of argument then, consciousness cannot be easily fitted into an objective scientific picture of the world: it remains a 'hard problem'. Francisco Varela proposed a way of proceeding that he claims dissolves rather than solves this problem. Rather than proposing any new theoretical framework, he advocates a particular research strategy – neurophenomenology - that he characterized as "a science of consciousness as if experience mattered" (Varela, 1998). Varela's starting point is the observation that any third person objective description arises as an invariant focus for a community of embodied, situated subjects endowed with conscious experience. This observation is usually either overlooked or overrated. It is overlooked, or regarded as trivial or parochial, by philosophers or scientists who think of invariance as just our way to discover reality that lies behind the superficial situated appearances. On the other hand it can be overrated by skeptics or relativists, who use it as a weapon against any claim of knowledge. Varela's central idea was that in the science of consciousness, one should neither try to absorb the subjective into the objective domain, nor objectivise the subjective, nor give the subjective any kind of supremacy over the objective. Rather, one should go

back to the experiential realm from which the very dichotomy between subjectivity and objectivity arises, and then establish within it a system of mutual constraints. In practice this involves a careful elaboration of first person statements in a process of phenomenological reduction, so that they can be properly related to third person statements. Varela, thereby, argued that science as a whole should no longer be restricted to describing structures that are invariant across a range of situations. It should be expanded to include both (a) regulated mutual relations between situated accounts, and (b) relations between situated accounts and their own invariants. As he put it science is extended so as to include a "dance" of mutual definition taking place between first-person and third-person accounts (Varela, 1998, 42).

Michel Bitbol (2002) endorses Varela's innovations and develops them by arguing more generally for a science in which situation matters beyond Varela's science of consciousness in which experience matters. In particular he points to physics, which is normally considered as the prototype of an exclusively objective science, as a case that requires an on-going dialectic between the objectified structures and a network of situated (actual or potential) subjects. In line with some of the approaches I discussed in chapter 3, Bitbol argues that many of the so-called 'paradoxes' of quantum mechanics can be dissolved if science is understood in Varela's wider sense. Thus, he argues that there exists a very detailed parallel between the mind-body problem and the measurement problem in quantum mechanics – "Both problems arose from an unavoidable blindspot in objective description. Both problems motivated a (fruitless) struggle aiming at encompassing the blindspot of actuality within the very objective structure that results from systematic elimination of situated actualities in favor of inter-situational invariants" (Bitbol, 2002, 212).

Bitbol takes this parallel to be highly significant, because it reveals the common limits of scientific knowledge classically conceived, and because it calls accordingly for a general redefinition of science. This redefinition would involve stretching the method of science in order to enforce a strong reciprocal relation between its objective contents on the one hand, and what

the very act of objectification forced one to leave out on the other hand. He envisages this 'participatory' science as "an ever-developing program of research about the neural correlates (or necessary conditions) of experiential contents, without bothering to look for an explanation of conscious experience by neural processes" (Bitbol. 2002, 213). Bitbol's concept of 'participatory science' seems consistent with Fuchs' 'participatory realism' interpretation of quantum mechanics that I discussed in chapter 3, section 4.

Bitbol argues that this redefinition of science involves an epistemological leap, which can only be compared with Darwinism. Before Darwin, natural science was methodologically restricted to reproducible states of affairs and lawlike necessity. Darwin encompassed contingency within the scientific domain by extending the methods of science to a natural history of random (genotypic and phenotypic) variations plus 'selection of the fittest'. Bitbol argues that Varela's innovation involves a similarly significant expansion of science to encompass subjectivity, or more generally situatedness. Bitbol understands the expansion of 'science' to 'participatory science' as returning it to closer to its original meaning as co-extensive with natural philosophy.

However, Claire Petitmengin (2017) characterises this approach as "a light or mild neurophenomenology". She argues that this cannot solve the hard problem of consciousness because it only involves correlation without getting at the causation. As Petitmengin put it -"it amounts to comparing neural and phenomenal *structures,* which does not shed any light on how lived experience may arise from cerebral activity". In contrast, she argues for a "deep or radical neurophenomenology … [that] aims at investigating the process of co-constitution of the subjective and objective poles, within lived experience". She claims to identify a deeply pre-reflective dimension of our subjective experience – what she calls the "felt" or "source" dimension, that is gestural and rhythmic, has precise transmodal sensorial submodalities, and seems to play an essential role in the process of emergence of all thought and understanding.

Petitmengin claims that this "felt" state is experienced without being conscious of a self; she understands it as the state that Buddhist meditation

practice calls “clear vision”. Thus she claims that radical neurophenomenology shows that the “*what it is like*” of experience is not necessarily *for a subject.*

However, the fact that Petitmengin identifies felt meaning with bodily experience suggests that there is still a basic element of subjectivity present. Petitmenigen does accept that even in everyday experience we do sometimes have access to this level of ‘felt’ experience – as for example in the discovery of a new idea. But, she claims, in the limit case subjectivity dissolves or vanishes. The fading of the feeling of individual identity might make sense in the same way that we can forget *who* we are when in are absorbed in some task or game. But even in these situations we never lose the feeling *that* we are.

Felt meaning (the feeling of being alive) is better understood in terms of our pre-reflective selfhood, which allows us to move seamlessly between the objective and subjective points of view. As Anna Ciaunica and Jane Charlton (2018) explain it, we shift back and forth across the bridge between the external realm of *Korper/Leben* (the body as a physical object with the capacity to survive) and the internal one of *Leib/Erleben* (the body as a subject that we live though from the inside). When it is functioning well, this experiential openness remains tacit and imperceptible, even as it pervades and sustains almost everything we do. But it is always the self’s presence *to us* that hooks us to the world. The self is the vital component of what it even means to experience reality.

5.7 THE METAPHYSICS OF BEING HERE

Although Bitbol advocates engaging in an expanded concept of science, he recognizes that any such engaging inevitably involves a “tacit yet creative presupposition”, or “ubiquitous background” – “the sense of mystery which arises from just ‘being there’, in this special human and individual situation” (Bitbol, 2002, 213).

We have knowledge of this existential and qualitative aspect of consciousness precisely because it is our way of being – we know it as the way that we find ourselves living. Being here we can know how the world, as what is in some ways separate from us, appears to us in a first personal way, giving us a subjective sense of our own existence. It cannot be like any content of experience because it is the way in which any particular content occurs. Objective science is an investigation of the contents of consciousness. A participatory science can include correlations between stabilized experiences and neural processes. However neither type of science can itself comprehensively investigate consciousness.

As Michel Bitol argues "Consciousness is and remains methodologically primary. Any objective descriptions arise, in history and on a day-to-day basis as well, as an invariant structural focus for subjects endowed with conscious experience" (Bitbol, 2008, 4).

Knowing-by-being is therefore prior to scientific knowledge, and thus imposes a constraint on any comprehensive theory of reality; any such theory must incorporate – and be consistent with – an account of ourselves, as we know ourselves. As Philip Goff (2017) argues, phenomenal consciousness is a hard datum that any adequate theory of reality must accommodate *unrevised*. He calls this –

'The Consciousness Constraint – Any adequate theory of reality must entail that at least some phenomenal concepts are satisfied" (Goff, 2017, 3).

He argues that this is a very significant constraint on metaphysical enquiry, "as it tells us that one of our ordinary pre-theoretical concepts *gets the world exactly right*." So while "I cannot rule out for certain" that "my entire experience of the external world is an illusion", "I can rule out the possibility that I have no inner life whatsoever". Thus Goff argues that the Consciousness Constraint has profound implications for scientific theorizing, setting strict limits on its potential for rewriting our conceptual scheme –

> "If it turns out that our scientific world view is inconsistent with the reality of inner experiencing – according to our ordinary notion of inner experiencing – then it is the scientific story rather than consciousness that must budge" (Goff, 2017, 4).

This completely reverses the order of priority specified by Ladyman and Ross that I mentioned in section 5.2. –

> "If cognitive science concludes that mental concepts do not track any real patterns then the theory of mind will have to go" (Ladyman and Ross, 2007, 254).

On Goff's view then, a comprehensive picture of reality, would have to be wider than any scientific picture of the world – it would have include what we know by being. This, then, is in line with Seager's fourth option of modifying metaphysics, which involves rejecting metaphysical scientism – the view that science (especially physics) is the sole source of knowledge about the ultimate nature of reality.

Giving up on metaphysical scientism does not necessarily imply anti-realism or a-realism about science. Science still investigates reality, but not all reality; science is situated within a wider human context. The scientistic claim is that science is the only reliable way of finding out about the world - "… science just *is* our set of institutional error filters for the job of discovering the objective character of the world … science respects no domain restrictions and will admit no epistemological rivals" (Ladyman and Ross, 2007, 28).

However, institutions, even scientific institutions are human creations. Repeated iterations of institutional error filters may lead to coherent results but hardly guarantee the epistemic supremacy of science. Physics may be sovereign in its own domain of mapping mathematical models of the behavioural dispositions of matter or fields, but there is more to reality than what can be captured in the mathematical vocabulary of physics.

As Philip Goff argues "consciousness is not something to be squeezed into the world we already know about from science"; rather it is better seen as "an *epistemic starting point,* on a par with the epistemic starting points we get from observations and experiments" (Goff, 2017, 11). On this view, consciousness is one of the best clues we have to working out what reality is really like. Frank, Gleiser and Thompson (2019) point to the 'entanglement' of our experience and what we call 'reality' –

"Our experience and what we call 'reality' are inextricable. Scientific knowledge is a self-correcting narrative made from the world and our experience of it evolving together".

And as Mary Midgley points out – "… the way in which the universe works isn't confined to the things that the sciences tell us about. That universe has hugely many aspects. It includes ourselves and our direct perceptions" (Midgley, 2014, 2).

Erwin Schrödinger, one of the founders of quantum mechanics was very aware of the limits of science –

> "… the scientific picture of the real world around me is deficient. It gives a lot of factual information, puts all our experiences in a magnificently consistent order, but it is ghastly silent about all and sundry that is really near to our heart, that really matters to us. It cannot tell us a word about red and blue, bitter and sweet, physical pain and physical delight; it knows nothing of beautiful and ugly, good and bad, God and eternity. Science sometimes pretends to answer questions in these domains, but the answers are very often so silly that we cannot be inclined to take them seriously" (Schrödinger, 1996, 95).

What we know by being is as much a part of the world as what we discover by science or any other human endeavour.

The success of science over the past few hundred years has been spectacular and has transformed human knowledge and human lives. However science remains a human achievement and has to start from the data that humans normally perceive. As Frank, Gleiser and Thompson (2019) put it –

"Science is essentially a highly refined form of human experience, based on our capacities to observe, act and communicate."

Thus, science is an important branch of human knowledge – but it is not the only one –

"Like other intellectual disciples – history, poetry, music, mathematics and the rest – science has grown out of the pre-existing social soil and is wholly dependent on it. Like them, it is a subculture" (Midgley, 2014, 2).

In this chapter I have completed my review of Seager's four 'escape routes' that I introduced in chapter three. They are possible escape routes from the paradox of consciousness that point to the difficulty of accounting for consciousness, as we know it by being conscious, in terms of the objective scientific view of the world. Using arguments from phenomenology, our first person knowing of ourselves, and biology, I favour Seager's fourth option that he labels 'modifying metaphysics' which involves a rejection of scientism but not of science. This is metaphysical in that scientism is the claim that science is our only access to reality, which is reached by abstracting away from our first person experience. Rejecting this involves realizing that our first person knowing of ourselves is the part of reality that we know with the greatest certainty. Science is a model-based method of investigating aspects of our world from within. It involves abstracting away from our lived reality in order to understand the world in various ways. But science on its own cannot give us a complete picture of reality.

In chapters three and five I have focused on consciousness which as I argued in chapter one is the initial primary fact about the way we find ourselves living. In the next chapter I return to the wider focus on selves that I introduced in chapters one and two, and developed in chapter four. It is selves or subjects that are conscious.

Are there selves? What in the world are selves? Could there be group selves? These are some of the questions I investigate in the next chapter by using the framework of counting selves.

CHAPTER 6. COUNTING SELVES

6.1 How many selves are there?
6.2. None
  6.2.1 Buddhist No-Self
  6.2.2 Relational Persons
  6.2.3 Illusionism
6.3. One Self
6.4 Group Selves
6.5 Elephants, Octopuses, Cells & Selves

6.1 HOW MANY SELVES ARE THERE?

A central theme of this dissertation is that selves are an important feature of 'the world as I found it'. This chapter aims to clarify the concept of self by considering some alternative views. I use the device of counting selves as a way of grouping these views and contrasting them with my view. In the last section and the following chapter I develop a more positive account of selves – not only can we count them, but they count – they are signifcant.

So, how many selves are there?

Here are four possible answers –

0

1

7.8 billion

More than 7.8 billion

The default position argued for by knowing-by-being is that each of us is the subject of first-person perspective, the experiencer of experiences and the source of actions. From the way we find ourselves existing we have a direct insight into what a self might be; it is generally acepted that we at least have a sense of being a self. This is a primary fact about the world as I found it. And, as I argued in Chapter 1, intersubjectivity is another primary fact about the way we find ourselves living; other humans are selves like us.

So an initial count of selves at 7.8 billion (the current human population of the world) seems reasonable. This gives us an initial grasp of the concept but further investigations are useful in clarifying the nature of this self. Is our initial grasp of the concept too promiscuous or too restrictive?

On the one hand, some argue that we are mistaken about our sense of self. According these views, our sense of self is an illusion or misunderstanding and that there is ultimatelty is no such thing. Others argue that while there are individual streams of experience the subject of these streams is one universal subjectivity. On the other hand, some argue for the wider application of selves to include some groups, and other or all living entities.

In this chapter then, I am using the framework of counting selves to consider some positions that may seem to contradict my concept of self or sugest that it is too restrictive. In the last section I use the discussion of the idea that all living entities are selves to clarify the significance of selves in the universe. This will lead on to the next chapter which argues that selves count (are important), because they have significance for, and add value to, the universe.

6.2 NONE

In this section I consider three ways to argue that our sense of self may be a mistake or illusion. Some Buddhists want to dissolve our sense of self into the flux of the universe, while some Confucians want to dissolve our individuality into our social roles; illusionists argue that the self is an illusion created by brain states.

6.2.1 BUDDHIST NO-SELF

While the view that there is no self has been proposed in western philosophy by some philosophers, notably David Hume (1739/1975) and Derek Parfit (1987), it has been more extensively explored by the Indian

Buddhist tradition, notably in the theory of non-self (*anātman*). Hume observed –

> "Whenever I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never catch myself at any time without a perception and never can observe anything but the perception" (Hume, 1739/1975, 252).

Like Hume, Buddhist philosophers point to the fleeting nature of all we find when we observe our inner states; nothing persists in the way that a self would presumably have to. They claim that the 'I' is posited to explain the felt unity among the inner states.

Differently from Hume however, the Buddhists have a particular motivation for, and method of investigation of, these questions. Their motivation is primarily to relieve suffering (*dukkha*); they see our ordinary sense of self, as the source of a certain sort of deep-seated suffering that results from our identification with an 'I' and our appropriation of things as 'mine'. Buddhist meditative practice aims to undermine all identification and appropriation. In support of this, Buddhist philosophy attempts to show that there is no entity that might serve as a referent of 'I', and to explain how the belief that there is such an entity comes about. Buddhist philosophical methodology is therefore primarily metaphysical. As Jay Garfield (2015, 2) explains –

"The origin of *dukkha* is in primal confusion about the fundamental nature of reality, and so its cure is at bottom a reorientation toward ontology and an awakening (*bodhi*) to the actual nature of existence."

Garfield argues that the Buddhist no-self doctrine is best understood as following from Buddhist metaphysics more generally. Buddhist metaphysics involves three central, tightly linked theses regarding the nature of reality: all phenomena are dependently originated, impermanent, and empty of intrinsic identity. As Garfield explains it, dependent origination is spelled out as a kind of brute regularity, which is taken to characterize reality quite generally. It guarantees that nothing is an independent existent. The only account we can give of anything is in terms of its relations to everything else.

Furthermore, all phenomena (with the exception of *nirvana*) are impermanent and their existence is regarded as momentary. At every moment, everything is changing, even if only in imperceptible, minute ways. Finally, all phenomena are empty of intrinsic identity - what we usually take to be things that endure over time are in fact continua of momentary, causally interacting events.

According to Buddhism, this metaphysical picture applies not only to objects but also to us. We are dependently originated, impermanent and empty of intrinsic identity: we are selfless persons. On the Buddhist view, the existence of a person consists in the existence of five *skandhas* (bundles or aggregates) organized in the right way. The five *skandhas* are *rūpa* (the body or corporeality), *vedanā* (affect and sensation), *samjñā* (perception and cognition), *samskāra* (conditioning and voilition), and *vijñāna* (consciousness). As with the person, these five *skandhas* are not to be taken as independent things: rather they are "interdependent aspects of a causally and functionally integrated psycho-physical system or process" (Mackenzie, 2011, 242).

Although all the major Buddhist schools say human beings are selfless persons, there is disagreement about the ontological implications of this. This disagreement turns on competing accounts of the concept of emptiness.

According to the Abhidharma school to be empty is to have no intrinsic nature or ontological independence and therefore, to be nothing more than a conceptual construct or convenient fiction. Conventionally, for practical purposes, it makes sense to ascribe identity to a life or an object over time, but ultimately, from this standpoint, nothing persists for more than a moment, and therefore the fiction of the self should be ultimately abandoned.

The Mādhyamikas agree that to be empty is to lack inherent existence (*svabhāva*), but also argue that this notion of inherent existence is itself untenable. On their view, positing ultimate, non-empty phenomena rules out the possibility of a coherent account of the conventional world. The Mādhyamikas, therefore, argue for not only the selflessness of persons, but also for the selflessness (emptiness) of all phenomena. But as Matthew

Mackenzie points out – "Within the Mādhyamika framework … for a thing to be empty is not for it to be unreal, but rather for its existence and nature necessarily to depend on other things … the self is empty, *just like everything else"* (Mackenzie, 2011, 258/9).

By ascription of conventional reality, but not ultimate reality, to what many Western philosophers would regard as unitary entities or processes, the Mādhyamakas take themselves to be taking to a middle path between nihilism (the denial that these things exist in *any sense*, or that they ground truth in *any sense*) and reification (the view that because they exist and can serve as truthmakers in *some sense*, that they exist ultimately).

Thus, the conventional/ultimate distinction ('two truths') is not as stark as it might first appear. As Garfield glossing Tsongkhapa, glossing Candrakirti puts it –

> "Even though Mādhyamikas argue that nothing exists truly—that is, ultimately, or independently—we must accept that things exist conventionally, for otherwise, we would not even be able to say that they do not exist ultimately. And when we say this, we do so truly, and so conventional existence, or conventional truth, must ground truth. If it couldn't, nothing could" (Garfield, 2015, 47).

As Garfield points out, the structure of this argument about existence and truth, is very similar to Wittgenstein's argument about knowledge in *On Certainty.* Wittgenstein (1969) argues that our epistemic conventions are just that, conventions; nonetheless, they underlie our justificatory and critical practices; if they could not, nothing could.

On this reading of the two truths doctrine (conventional and ultimate truth), Buddhist metaphysics of the self is not so much about denying our existence; rather it is about explaining the way in which we exist. The claim is that we as persons exist dependently on everything else: not independently of everything else. This is consistent with the primary facts that I outlined in chapter one: while we are in some way independent, we live with others as part of a world that affects us and is affected by us. In terms of western philosophy this might be equivalent to the claim that, whatever we are, we are part of the natural world and not Cartesian souls.

Also, the ontological claims of Mādhyamikas Buddhism fit well with western enactive views that understand entities in terms of stable processes. As Evan Thompson explains it –

> "Phenomena at all scales are not [independent] entities or substances but relatively stable processes, and since processes achieve stability at different levels of complexity, while still interacting with processes at other levels, all are equally real and none has absolute ontological primacy" (Thompson, 2007, 441).

The no-self view in Buddhism is a response to Brahmanic and Vedantic views that posit an immutable self – an *ātman*. But as Owen Flanagan writes

> - "Even if there is no ātman backing you up. You—the person you are—are still there. Furthermore, you have a certain constancy, a personality—a character even. Indeed you must be something rather than nothing, you must have an abiding presence—not, of course, one of self-identicality over time but of personal continuity and connectedness—for the recommendations to live wisely and virtuously to make any sense" (Flanagan, 2011, 134).

Having 'a personality', 'character' and 'an abiding presence … of personal continuity and connectedness' might be, at least in western parlance, sufficient to justify the use of term 'self'. However, if Buddhists prefer to use the term 'person', so be it.

Overall, Buddhist metaphysical claims that everything exists dependently on everything else, as part of the flux, is consistent with western ideas that everything is part of nature. And the claim of 'no-self' can be understood as a claim that we are not supernatural beings; whatever we are, we are part of nature - like everything else. This is consistent with the kind of self that I argue we know by being.

6.2.2 RELATIONAL PERSONS

Another approach that appears to deny the significance of individual selves is the Chinese idea of society and relational persons. While Chinese thought is predominantly human-centred, it is practice orientated. As Roel Sterckx puts it –

> "The great questions that have occupied China's brightest minds are not about who or what we are, but rather about how we should live our lives, how we relate to others, how we should organize society and how we can secure the well-being of those who live with us and for whom we are responsible" (Sterckx, 2019, xi).

Thus Chinese thought tends to stress the primacy of what is considered good for the society over any idea of individuality to such an extent that they find it unnatural to refer to themselves by the first person singular pronoun. Here is how Xialou Guo (2017) describes her experience of how she had to change her way of speaking and thinking when she moved to the west –

" … how could someone who had grown up in a collective society get used to using the first-person singular all the time? The habitual use of "I" requires thinking of yourself as a separate entity in a society of separate entities. But in China no one is a separate entity: either you were born to a non-political peasant household or to a Communist party household."

The importance of the family, and society more generally, over the individual is deeply rooted in Chinese culture generally and particularly in Ruism or Confucianism. According to the central Confucian virtue of filial piety even one's body (or physical existence) is given by your parents and nourished by the family, so that self-preservation or self-development is not seen as a matter of self-interest; rather it is seen as an obligation to take care of the family 'property'. As Vittorio Bufacchi and Xiao Ouyang (2017) point out "the idea of the 'self' is derogatory in China, due to its association with the 'private' and 'self-interest'". Also, Confucian ethics is expressed in terms of various concrete social roles or relationships within specific contexts, rather than in terms of an individual moral agent. On this view then we are not the *same* person throughout our lives. Rather, Confucians talk of persons

in terms of *changing* and *growing.* As Roger Ames and Henry Rosemont (1998, 27) explain it "… growing up through the cultivation and articulation of relationships makes us *different* persons, as does getting married, becoming a parent, and so forth. As the quality and quantity of our relationships proliferate, so, too, are we extended in the world". This idea of a person being extended in the world is an interesting idea that parallels in some ways, the recent western ideas of an extended mind.

In his 2015 book, *Against Individualism,* Henry Rosemont aims "to show that the view of human beings as most fundamentally free and rational, autonomous individual selves is almost certainly false, and more than that, mischievous: its celebration and defence of freedom comes at the expense of social justice, and peace." He considers "the views of the early Confucians on what it is to be a role-bearing person … suitably modified for contemporary circumstances and sensibilities … [as] a genuinely viable alternative answer to the question of what it is to be a human being" (Rosemont, 2015, xii). As Rosemont explains it, the Confucian idea of a role-bearing person is that persons are completely constituted by the social roles they occupy; there is no self that could be separated from the person's social roles.

Rosemont argues that the idea of the individual self, based on self-awareness that entails rationality, is one of the most deeply rooted constructs in the history of Western intellectual history. Its modern form is expressed by John Locke as follows - "The freedom then of man, and liberty of acting according to his own will, is grounded in his having reason, which is able to instruct him in that law he is able to govern himself by… Men being, as has been said, by nature, all free, equal and independent …" (Locke, 1689/1986, 36). And as Niklas Luhmann points out – "Sociological theory … conceives history as a process of increasing individualism" (Luhmann, 1986, 313).

Rosemont exemplifies the contrast between the Chinese and Western concepts of human beings with the following two quotations –

- Mencius – "No Human being is insensitive to the suffering of others. If a person sees a child on the verge of falling into a well, he would surely be moved to compassion, not because he wanted to get in the good graces of

the parents, nor because he sought the praises of his neighbours. … A person insensitive to the suffering of others is not human." (Mencius, 2A6)
- Kant – " (T)he unsocial sociability of men .. bound together with a mutual opposition which constantly threatens to break up the society … he has a strong propensity to isolate himself from others, … expects opposition on all sides because, on knowing himself, … he knows he is inclined to oppose others" (Kant, 1784/1981, 15-16).

Rosemont argues that while Mencius "suggests strongly that we are not isolates, that our instincts toward one another are basically social, and need to be *nurtured",* Kant suggests that "our instincts are strongly selfish and anti-social, but the through the use of reason they can and must be *controlled"* (Rosemont, 2015, 35).

Rosemont says that according to Kant, the self-consciousness that attends the use of the first person pronoun 'I' is a defining assumption that elevates the human being above all other forms of life by giving them a sense of persistent identity and control. But Rosemont argues that the personal pronoun 'I' alone does not offer us either a concept of something, or of something else. Rather, "it is a linguistic entity – an indexical – that refers us to a particular situation like 'now', 'here', 'that' within which the 'I' utterance takes place" (Rosemont, 2015, 39).

Arguing for the relational view, Rosemont says that in everyday life, the question 'who am I?' is answered in relations to others – a son, a husband, a father, a colleague, a teacher etc. And "(w)hen all of these interrelationships have been specified, and their interrelations spelled out … then I have been altogether individuated, but with nothing left over with which to piece together an individual self; or so it seems to me" (Rosemont, 2015, 49). He calls the belief in an essential individual self, residing in each of us, "an ontological fiction" (Rosemont, 2015, 50).

As David Hall and Roger Ames (1987, 153) explain it – "The very notion of 'individual' in the strictest sense is suspect from the Confucian perspective. 'Sociality' is the very root of existence." On this view then, Chinese culture may not have the conceptual apparatus, or see the need, to distinguish the 'community' from the 'individual'. This makes it difficult, if

not impossible, to straightforwardly apply western ideas such as individual human rights to China. In the Chinese worldview, 'human' is understood as a collective rather than an individual term. As Herbert Fingarette (1983) put it – "For Confucius, unless there are at least two human beings, there are no human beings."

On this view, social stability will always be more important than individual freedom. The economic success in lifting millions of people out of poverty and the public health success in controlling the spread of the coronavirus, outweigh any individual human rights violations that occur in their achievement.

However, although this view may seem to totally dissolve any notion of individuality into one's relationship with others, I think we can still identify a self at least in a minimum form as a subject of experience in the Chinese tradition. The emphasis on the necessity for cultivation and becoming an exemplary person, are clearly aimed at teaching individuals how become better (less private/ less self-interested) persons. An obligation, even an obligation to look after your body as family property, is an obligation of an individual.

The Chinese emphasis is on how you should behave rather than what exists, or, 'can you eat it?' rather than 'what is it'. As the writer Lin Yutang (2001, 47) partly satirically quipped the Chinese did not pursue the biological sciences because they reduced what is knowable to what is edible – "… the reason why the Chinese failed to develop botany and zoology is that the Chinese scholar cannot stare coldly and unemotionally at a fish without thinking immediately of how it tastes in the mouth and wanting to eat it." This emphasis may mask the subject who should behave or the person who must eat, but there is still a subject of experience there, at least implicitly.

On the relational view however, "there is no reference to some core human *being* as the site of who we *really* are and that remains once the particular layers of family and community relations are peeled away" (Ames, 2011, 122).

But Rosemont asks rhetorically - *"Who*, exactly, bears these Confucian roles?" … In the end, the best answer to the question is probably simply 'The body'….

It is not clear to me that anything more than my name and its embodied denotation is needed to answer such questions" (Rosemont, 2015, 52). He also uses the phrase "role-bearing persons" because "we are basically *constituted* by the role we live in the midst of others" (Rosemont, 2015, 95).

In explaining how role-bearing persons live their lives however, Rosemont makes the following observation – "The *ren dao,* ('Way of Humankind') however, is not singular. Each of us is unique, and hence we must each forge our own path through life with our fellows on the basis of our abilities and personal qualities, which distinguish our interactions with others from the interactions of other similarly unique role bearers" (Rosemont, 2015, 93). As Confucius (*Analects,* 15.29) himself put it – "It is the person who is able to broaden the way (*dao*), not the way that broadens the person" (Ames and Rosemont, 1998, 190).

This description of each of us as unique role-bearers, with abilities and personal qualities, seems to me to be sufficient to justify at the use of the term 'self'. And in fact, some Confucian authors do specifically refer to unique individual selves as the foundation for the family and society. According to Mencius (4A.5) –

"The Empire has its basis in the state, the state in the family, and the family in one's own self" (as translated by D.C.Lau in 1970).

Or -

"The foundation of the world lies with the state; the foundation of the state lies with the family; the foundation of the family lies with the self" (as translated by Irene Bloom in 2009).

This is in line with *The Great Learning (Daxue),* which was originally a chapter in *The Book of Rites,* one of the five pre-Confucian classics. This chapter was adapted, commentary was added, and it became one of the Four Books (the others being – *The Analects, The Doctrine of the Mean, & The Mencius*) that came to be regarded as distilling the essence of Confucianism. *The Great Learning* is a short text that was memorized by generations of students, and used in the Chinese state examinations up until 1905. Using the metaphor of a tree, it describes the successive stages through which social order is established: one starts by cultivating the individual self ('the

root'), then extends this to the family, and finally to the state and the world at large ('the branches') –

"Only after knowledge is brought about will your intentions be sincere. Only after your intentions are sincere will your heart be rectified. Only after your heart is rectified can you cultivate your own person. Only after your own person is cultivated will your family be regulated. Only after your family is regulated will the state be ordered. Only after the state is ordered will the world be at peace" (Translated by Plaks, 2003).

As Roel Sterckx explains it –

"The point *The Great Learning* makes is that in order to look outwards, you need to be prepared to look inward and focus on the inner self: being true to the self leads to good inter-personal relationships within the family and, via the family, with the state" (Sterckx, 2019, 218).

On this reading then the inner individual self is the 'root' of society - not an 'ontological fiction' as Rosemont suggests; and this concept of individual selves can be found throughout Chinese thought.

Also, at least on some Western conceptions, the self is not an isolated individual; rather, as I suggest from the way we find ourselves living, the self is always with others in the world. Rosemont's overall project is ethical and political. He chooses a particular Western notion of isolated rational self as a contrast to his Confucian inspired role-bearing persons. The Confucian ideas of reciprocity and roles have undoubtedly much to say to Western ethical and political ideas. It is not clear however, that this needs to be necessarily accompanied by a denial of the reality of selves. The question of what it is to be a human being needs to be answered using the resources of both, and other, traditions.

6.2.3 ILLUSIONISM

Another way of arguing that there are no selves is to contend that our sense of self or our consciousness is an illusion. I argue in chapter one that consciousness is the primary fact that we know about ourselves, and our only access to ourselves, and the world. If consciousness is an illusion we cannot know that we are a self. Illusionism about selves and consciousness is therefore closely intertwined.

It is a 'hard problem' for physicalist philosophers to explain how there can be phenomenal consciousness and selves who are phenomenally consciousness. We still have no explanation of how the conscious selves that we experience ourselves to be, could have evolved from a purely physical universe.

One suggested way of avoiding this hard problem is to argue that consciousness and even selves are illusions or simulations. So for example, Jay Garfield (2016) argues that "There is no phenomenal consciousness; there is nothing 'that its like' to be me… the sense that there is such a kind of consciousness is an instance of cognitive illusion."

According to illusionists, our sense that it is like something to have conscious experiences is a result of the fact that we systematically misrepresent them. On this view then the task for a theory of consciousness is to explain our illusory representations of phenomenality, not phenomenality itself, and the hard problem is replaced by the illusion problem (Frankish, 2016, 11).

Dan Dennett (2016) stresses that what illusionists are concerned with is explaining how the illusion of phenomenal consciousness comes about – *"how the magic is done".* As Dennett clarifies it, the magic is *stage* magic with physical causes not *real* magic with supernatural causes –

"In today's world, if not in the Dark Ages or even the Renaissance, the standard default assumption about any feat of stage magic we encounter is that it is (somehow) the product of everyday physical causes, involving perhaps magnets and electricity, or even holograms, but not psychokinesis,

clairvoyance, or the assistance of any poltergeists, goblins, or other supernatural agents" (Dennett, 2016, 66).

Frankish (2016, 12-14) identifies three broad options for explaining phenomenal consciousness –

- radical realism – which treats phenomenal consciousness as real and inexplicable without radical theoretical innovation.
- conservative realism – which accepts the reality of phenomenal consciousness but seeks to explain it in physical terms, using the resources of contemporary cognitive science or modest extensions of it.
- illusionism – which shares radical realism's emphasis on the anomalousness of phenomenal consciousness and conservative realism's rejection of radical theoretical innovation. It reconciles these commitments by treating phenomenal properties as illusory.

Frankish and Dennett therefore argue that illusionism should be the obvious default theory of consciousness, in that it claims to conservatively explain phenomenal consciousness, as it appears - as an illusion engendered by nature (Dennett, 2016, 67).

Dennett compares consciousness to the user illusions created by graphical interfaces (Dennett, 1991, 216-20) or apps (Dennett, 2017, 341) by means of which we control our computers –

"The user interface of an app exists in order to make the competence accessible to users – people – who can't know, and don't need to know, the intricate details of how it works. The user-illusions of all the apps stored in our brains exist *for the same reason:* they make our competences (somewhat) accessible to users …" (Dennett, 2017, 341).

On Dennett's view then, phenomenal consciousness presents us with metaphorical representations of real neural events, which facilitate certain kinds of mental self-manipulation but give us no deep insight into the processes involved. Dennett points out that we, and other animals, can perform many skillful actions with only a vague conception of what we are up to. Humans, uniquely, however, often retrospectively attribute reasons for our actions. We achieve this by using 'cultural-borne memes' of self-justification, self-reproach and self-criticism, which enable us to plan ahead

and design our own future acts. Dennett argues that our ability to do this kind of thinking is not accomplished by any dedicated brain structure not found in other animals. As he puts it, there is no 'explainer-nucleus'. Rather, "our thinking is enabled by the installation of a virtual machine made of virtual machines made of virtual machines" (Dennett, 2017, 341).

Dennett explains how we developed these self-monitoring systems in terms of communication. He quotes David McFarland (1989) – "Communication is the only behavior that requires an organism to self-monitor its own control system." Thus Dennett argues that communication requires a central clearinghouse of sorts in order to buffer the organism from revealing too much about its current state to competitive organisms (Dennett, 2017, 342).

He points out that all organisms, from single cells to elephants, have a rudimentary 'sense of self', in that they need to keep track of which limbs are theirs and what they are doing with them. However, as we humans learned to communicate we needed to develop a rapport with each other and this required perceiving *ourselves* in carrying out these behaviors, so that we could keep track of which thoughts are ours and whether we should share them with others. Dennett says that this is what gives us a more 'selfy' rather than rudimentary sense of self (Dennett, 2017, 344).

For Dennett however, this 'self' is no more than a "center of narrative gravity" (Dennett, 1991). Just as the center of gravity is an explanatorily useful theoretical fiction in physics, so Dennett argues, our self is a useful fictional character in a story written by a program in our brain. We only need, and therefore have, our stream of consciousness insofar as it gives us an edited digest of current activities that is available for communication with others. Thus he argues that the self is "not a dedicated portion of neural circuitry but rather like the end-user of an operating system" (Dennett, 2017, 345).

On Dennett's view then our *first-person* point of view of our own minds is not so different from our *second-person* point of view of others' minds – " we don't see, or hear, or feel, the complicated neural machinery churning away in our brains but have to settle for an interpreted, digested version, a user-

illusion that is so familiar to us that we take it not just for reality but also for the most indubitable and intimately known reality of all. That's what it is like to be us. We learn about others from hearing or reading what they say to us, and that's how we learn about ourselves as well" (Dennett, 2017, 345).

The way Dennett uses of the analogy of our minds with apps seems inconsistent. When he is talking about smartphone apps he refers to it as a user *interface;* but when he is talking about our minds he refers to a user-*illusion.* As he points out apps enable us to do things with smartphones without knowing the intricate details of how it works. However, there is no illusion here. An app is a good (real) app in so far as allows us to do something. The fact that we don't know the details of how it is implemented doesn't make it an illusion. Similarly, it might be the case that because our minds have to control the human being as a whole, they have to operate at this level rather than at the level of neural circuitry. This does not necessarily mean that our selves are illusory.

Dennett might be right about the evolutionary story about how our first person perspective came about. Lynne Rudder Baker (2013) among others supports the idea that second person perspective developed first with our acquisition of language. However, this fact need not necessarily contradict the current primacy of first person perspective.

Dennett's claim is that the unity implicit in the notion of the thinking-perceiving subject (or self) dissolves on close inspection as the mind separates into a collection of autonomous subsystems. Jenann Ismael (2006) however, argues that the Joycean Machine of consciousness plays an important unifying role, recombining informational streams by mapping them into an internal reconstruction of the external environment, and this is what allows deliberation to take place.

Dennett argues that the phenomenally conscious self is an illusion because he cannot identify any corresponding 'dedicated portion of neural circuitry' or 'explainer-circuitry'. Thomas Metzinger makes a similar argument –

"… there is no such thing as a self. Contrary to what most people believe, nobody has ever *been* or *had* a self. […] to the best of our current knowledge

there is no thing, no indivisible entity, that is us, neither in the brain nor in some metaphysical realm beyond this world" (Metzinger, 2009, 1).

What Dennett and Metzinger are arguing against here, however, is the claim that there is *an entity* called 'the Self' that is somehow distinguishable from the person as a whole; as Metzinger puts it, there is "no little man inside the head" (Metzinger, 2009, 208). Jan Slaby and Jan-Christoph Heilinger point out however, that most philosophers do not now argue that the self is a separate substantial entity. Rather, they point to examples of how selfhood might be construed as "a fluid process of reiterated agent-environment interactions in a formative social context, or as a matter of social attribution taken up and internalized by reflectively conscious agents, and so on" (Slaby and Heilinger, 2013, 87).

Metzinger claims that what we perceive as the world and as 'ourselves', is in fact a simulation generated by the brain – what he calls a projection into 'phenospace' –

"First, our brains generate a world-simulation, so perfect that we do not recognize it as an image in our minds. Then, they generate an inner image of ourselves as a whole" (Metzinger, 2009, 7).

Dan Zahavi labels Metzinger's position as "neural representationalism", according to which "we" are living in a vivid hallucination generated by the brain, and are never in direct epistemic contact with the world surrounding us (Zahavi, 2005, 16).

But Metzinger claims to know that reality is just physical –

"The world is just as your physics teacher in high school told you: an ocean of electromagnetic radiation, a wild and raging mixture of different wavelengths" (Metzinger, 2009, 20).

Slaby and Heilinger point out that Metzinger's theory is incoherent in that it places us in a position that, if we in fact were in it, we could not have any knowledge about. They argue that Metzinger exhibits a kind of blatant scientism –an exaggerated confidence in the epistemic power of the sciences. As they argue – "Given the fact that science obviously is a human endeavour and, as such, at any time tied to human epistemic capacities, it is highly doubtful whether science would be able to operate from such a privileged

standpoint, *fundamentally* different from and superior to all and every nonscientific epistemic positionings" (Slaby and Heilinger, 2013, 89).

An obvious objection to illusionism is that an illusion presupposes an audience. Illusionism therefore appears to be incoherent because it denies the reality of the audience it presupposes. The basic illusionist claim is that introspection delivers a partial, distorted view of our experiences, misrepresenting complex physical features as simple phenomenal ones (Frankish, 2016, 18). However although they deny the existence of experiences with phenomenal properties, they do not deny the existence of a form of consciousness (access consciousness), which consists in the possession of states with quasi-phenomenal properties. This maybe sufficient to allow for an inner *representer* of some kind – as Frankish explains it, this inner representer might consist of "a system, or set of processes, which generates introspective representations of sensory states. But this need not amount to an observer, still less a conscious one" (Frankish, 2016, 34).

He argues that if we do not need an inner observer to appreciate perceptual representations, why should we need one to appreciate introspective ones? Dennett (1991) argues, that once the brain has made discrimination, there is no need for another brain system to remake it, and all the work of appreciation and reaction can be (and ultimately must be) distributed among numerous unintelligent subsystems. Similarly, once an introspective representation has been generated, the work of reacting to it – of being impressed by the illusion – can, and must, be distributed across such subsystems. There need be no unified audience for the illusion smaller than the organism as a whole (or at least its central nervous system) (Frankish, 2016, 34).

However, Ned Block rejects the idea that creatures without phenomenal consciousness can have access consciousness (Block, 2007, 219-20,282). He agrees with Tyler Burge's (1997) claim that phenomenal consciousness is the 'core notion' of consciousness, and that access consciousness is 'parasitic' on phenomenal consciousness in the way that a parquet floor is parasitic on the sub-floor beneath it. Block argues that when we inquire whether a

creature is conscious, phenomenality is a crucial part of what we are inquiring about. Following this line of thought, Overgaard and Salice (2019) argue that conscious accessing cannot be made intelligible in terms of access consciousness alone. They argue that access consciousness to a belief (which is generally understood as 'being broadcast or poised for use'), is not sufficient to ensure that the belief is consciously accessed. Rather, they argue that conscious accessing of beliefs necessarily involves some phenomenal consciousness. Content that is "poised for accessing" will not be "*accessed"* without a "P- and A- conscious judgment" (Overgaard and Salice, 2019). If this is correct, it is not clear that Frankish's 'inner representer' could be the audience of the illusion of phenomenal consciousness – unless it was phenomenally conscious itself.

Evan Thompson (2015) argues that the self is enacted by self-specifying and self-designating processes. According to this way of thinking, the self isn't an independent thing or entity; it's a process. But this is different from saying that there really is no self or that the appearance of the self is nothing but an illusion.

According to no-self views, an individual stream of consciousness mentally represents itself as belonging to a self, but in reality no self exists, so the mental impression of there being a self is an illusion. Here "self" means an independent "I" or ego that's wholly present at each moment and owns the experiences that make up the stream of consciousness.

Thompson argues that although our sense of self is mentally constructed, it does not follow that there is no self or that the appearance of the self is nothing but an illusion. "Although some illusions are constructions, not all constructions are illusions. The self is a case in point. To say that the sense of self is a mental construction – or rather that it's a process under constant mental and bodily construction – doesn't logically imply that there is no self or that the sense of self presents an illusion" (Thompson, 2015, 359).

Thompson compares the self to an image in a mirror -

> "The image depends for its existence on the mirror – the mirror is the basis for the image – but the image isn't one and the same thing as the

> mirror, nor is it made of the same stuff as the mirror, for as an image it exists only in relation to an observer. Notice that the mirror image, though observer-dependent, isn't a subjective illusion. So too the self, though mind-dependent, isn't a subjective illusion" (Thompson, 2015, 364-365).

Nevertheless, Thompson says that the way that the self appears can involve an illusion, even if it's not the case that there is no self or that the appearance of the self is nothing but an illusion. "The illusion- or delusion – is taking the self to have an independent existence, like taking the mirror image to be really in the mirror. Notice that the image as such isn't an illusion: it's the taking of the image to exist in the mirror that's the illusion. Similarly, it's not the appearance of the self as such that's the illusion; it's taking the self to exist independently that's the illusion" (Thompson, 2015, 365).

As we have seen, Dennett argues that the phenomenally conscious self is an illusion because he cannot identify any corresponding 'dedicated portion of neural circuitry' or 'explainer circuitry', and Metzinger can find no such thing as a self. Both are arguing from a physicalist position. James Tartaglia (2016) takes this to be the central message of Keith Frankish's article on Illusionism –

"if physicalism is true, phenomenal consciousness must be an illusion."

However, Tartaglia argues that by rendering consciousness illusory for the sake of our physical conception of the world, the physicalist forgets where that conception came from –

> "With the physical conception in place, physicalist illusionism makes sense. But human beings built up that conception by working out what their conscious experiences were telling them about the world. If fictions were integral to our reasoning, we should never have believed the physical conception that tells us they are fictions."

The above arguments against illusionism about consciousness and selves are self-referential or self-reflective. Martine Nida-Rümelin (2016) applies this way of arguing most directly by showing that illusionism is itself an illusion. As she argues a central thesis of Frankish's argument for

illusionism is that illusionism is possibly true. However if illusionism is true, the illusionist cannot be understood as seriously accepting his own theory – "The (supposed) belief in illusionism thus turns out to be an illusion".

6.3 ONE SELF

The apparantly counterintuititive idea that there is only one self or one subject of all experience is an idea found in the Indian Advaita Vedanta tradition and more generally in mysticism. Erwin Schrödinger expresses the basic conviction of Vedanta as follows –

> "it is not possible that this unity in knowledge, feeling and choice which you call *your own* should have sprung into being from nothingness at a given moment not so long ago; rather, this knowledge, feeling and choice are essentially eternal and unchangeably and numerically *one* in all men, nay in all sentient beings" (Schrödinger 1964, 20 f.).

Wolfgang Fasching (2016) adopts this point of view and thereby argues that 'I' of subjectivity is non-plural – there is just one. Fasching's specific claim is that the ultimate subject of experience is not the individual but rather something that knows no multiplicity - 'pure awareness'.

Fasching arrives at this conclusion from his consideration of the special aspects of 'I'. As we saw in Chapter 2, Fasching argues that there seems to be a strange contingency about the fact that it is *me* who is this particular person. The fact of my existence – for me, from a first person perspective – seems to be a further fact beyond all the objective facts about the world. Fasching argues from this that my experiences are my experiences, not only because they have such and such experiential qualities, transpire at this and that particular time etc., but because they take place 'in me', in the subjectivity-space as which I exist. Thus the 'I' is not one of the experiences or the whole of somehow united experiences or some structure they assume, but rather the 'where' of their taking place. So he concludes that the I is "not so much ontologically located at the level of the individual things and events in space and time, but at that of space and time themselves: it has the

character of a dimension. It is a plane of being, and not simply one of the things that are inner-dimensionally located."

On this view then Fasching argues that the I is not an object in the world but rather the for-whom of all objectivity, the where of the stretching-out of the objective world itself - "A where which is not a place within the world but the 'primal place' of the world itself with all its 'places': the dimension of world-manifestation and therefore not locatable within the world manifesting to me". As Zeno Vendler put it - 'the empirical world, as perceived by a subject, points beyond itself, in spite of the fact that it is perceived as complete' (Vendler, 1984, 105).

But if the world not only appears to me, but also to others, is there not therefore, a plurality of such radically non-objective 'primal places'? Fasching says this is "an obvious absurdity". So "ultimately the idea of a plurality of Is is, thought out, simply unintelligible – thus we are lead to the 'Vedantic' thought: maybe there are no 'other Is', maybe all experiences which have ever been experienced were experienced by one and the same ultimate I which is me (just like everyone else). I am not an I but simply the I".

The crucial move that Fasching makes is from the observation that 'I' involves all my experiences taking place 'in me', to concluding that 'I' is the 'where' of my experiences taking place. This involves a shift from the individual 'me' to the universal 'where', which he can then claim, has the character of a 'dimension'. I argue that to remain true to the phenomenal experience of 'me', the 'where' is always 'here' – I always find myself located at a particular place and time – the 'where' at which my experiences take place is always 'here' in me not in a universal dimension.

Fasching doesn't deny that there are multiple streams of experience, but there is also an experiencing I to whom the streaming experiences are present – and his suggestion is that the plurality lies in the streams alone while the later, the to-whom as such, is strictly singular. He argues that it is clear enough what distinguishes the experiences of this stream from the experiences of that stream; yet it is by far less clear what should distinguish the to-whom of all those streaming experiences. However, this separation of

the stream of experience from the experiencer seems to be contrary to what he argued previously about 'I' and consciousness – "'I' and 'my consciousness' are not to be separated here: the I has its being in being consciousness (i.e. the experiencer has its being in experiencing) and in this sense is nothing beyond consciousness." However, identifying the 'I' with a universal feature such as 'pure awareness' seems to involve a separation of 'me' and 'my consciousness'.

The idea that there is just one universal consciousness is also the starting point for Itay Shani's theory of cosmopsychism (Shani, 2015). He proposes this theory as a holistic alternative to atomistic panpsychism, and as a general perspective on the metaphysics of consciousness. Cosmopsychism proposes that an omnipresent cosmic consciousness is the single ontological ultimate, which serves as the ground for the emergence of individual conscious creatures. While cosmopsychism is broadly a panpsychist theory of mind, nevertheless in its holistic commitments it functions as an alternative to the atomistic thinking, which dominates work on panpsychism in contemporary analytic philosophy. It is therefore claimed to be better able to deal with some of the problems associated with panpsychism notably the subject combination problem. It does this by arguing that there is only one ultimate, absolute cosmic consciousness, and seeks to explain familiar macro-level consciousness as being rooted in such cosmic origins. However, even if this does solve the combination problem, it creates a corresponding division problem of explaining how the single cosmic consiousness is experienced by each of us individually.

Cosmopsychisim can be seen as a mirror image of physicalism - a universal subjective theory rather than a universal objective theory. Like physicalism, it claims to be a universal theory of all there is. Both theories, I claim, ignore or seek to dissolve the source of any human theory whatsoever – our individual knowing of ourselves.

6.4 GROUP SELVES

Some recent literature on groups suggests that some groups can have features, or even be the kind of entity, that individuals are traditionally thought to have, or be. Margret Gilbert argues that "a group can in principle believe that *p* even though everyone in it personally believes the opposite" (Gilbert, 1987, 201), and that this collective belief "picks out an important, real phenomenon" (ibid, 200). Similarly, Deborah Tollefsen argues that some "groups really do have attitudes" that amount to "real patterns in the world" (Tollefsen, 2015, 111). Bryce Huebner (2013) claims there are rare cases of collective mentality in our world – groups that are minded in the same sense that we are. Naomi Ellemers (2012) argues that there can even be a group self that can become more important than individual selves in determining thoughts, feelings and behaviours. And Philip Pettit (2010) argues that some groups not only have "minds of their own", but also are "subjects in their own right".

According to Pettit, some groups, if they are organized in a suitable way, are subjects in their own right and therefore can have intentions and act independently of their members. The claim that it is rational for some groups to hold beliefs that the majority of its members do not hold is an insight secured from discussion of the 'discursive dilemma', that a multi-judge court may face in deciding a multi-premise judgement. If the court aggregates the judges' opinions about the individual premises (rather than aggregating the conclusions of each judge), then the court may deliver a judgement that is not held by a majority (or maybe not even any) of its members. The claim is that the judges in such a court would constitute a group and that therefore it is argued that certain groups can have beliefs (over and above the beliefs of individuals).

Joel Walmsley (2015) agrees with Pettit's analysis and takes the discursive dilemma to be a case of emergence, in C.D. Broad's (1925) sense. He therefore considers it to be a genuine example of a non-reductive group judgement. However the discursive dilemma takes a very narrow view of what is the relevant group. Judges in a court are paid employees who are

carrying out a particular institutional role in the legal system of a country. The choice of which decision procedure to adopt is always, at least potentially, a decision of the individuals of the wider society in which this particular discursive institution occurs. In the case of a court's judgement, the decision on how to aggregate the individual judges decisions will normally be specified by the constitution, which is put in place by members of the society. If no method of aggregation is specified, then the judges themselves would have to adopt a procedure – but this would always be subject to the potential or tacit acceptance, or not, by the wider community that the judges serve. In Pettit's other example - the case of employees of a company that they own themselves - they would have to come to an agreement about the aggregation procedure prior to its being applied and this would have to have the at least tacit acceptance of all concerned. In neither case however will the decision on how to aggregate the judgements or the judgements themselves, be decisions taken by any entity over and above the relevant members of the group in question.

Also I think Pettit doesn't take proper account of the way the purpose of a group may explain particular group decisions. He does recognise that it is purposive groups that "will almost inevitably confront examples of the discursive dilemma and that, short of deception, they will be under enormous pressure to collectivize reason" (Pettit, 2010, 176). And following Stoljar (1973) he recognizes that there are two types of purposive groups. Firstly, groups charged by an external authority to perform a specific function (such as museums, libraries, trusts and states and also more episodic groups such as appointment committees, juries or commissions of enquiry); and secondly groups that are associated with a characteristic goal, involving the outside world, the groups own members or a mix of both (such as a political party, a trade union or a group of friends planning a holiday).

In either case it is the purpose of the group, and therefore, the individuals whose purpose it is, that drives the pressure to collectivise reason. Groups that collectivise reason do so in furtherance of the purpose for which they exist. And this is the purpose of individuals who set it up or the individuals for whom it now performs a function. This purpose is external to the group

as such. It is not the group itself that has a mind - rather the group structure is the expression of the purpose of the group's founders, members and supporters. The fact that some groups collectivise reason therefore does not involve a dilemma and does not warrant them as being regarded as subjects in their own right.

Pettit seems to be ambivalent on this point. On the one hand he argues that "groups that collectivise reason deserve ontological recognition as intentional and personal subjects" (Pettit, 2010, 175). On the other hand he makes a mere supervenience type claim – groups supervene on individuals but there is no emergence.

> "I am not postulating any ontological mystery. The argument is consistent with the supervenience claim that if we replicate how things are with and between individuals in a collectivity – in particular, their individual judgments and their individual dispositions to accept a certain procedure – then we will replicate all the collective judgments and intentions that the group makes. Collective judgments and intentions may be discontinuous with what happens at the individual level but they cannot vary independently of what happens there; they do not constitute an ontologically emergent realm" (Pettit, 2010, 184).

So he is saying that while some groups deserve "ontological recognition' they do not constitute an "ontologically emergent realm". This contrasts with his title – "Groups with minds of their own" which he describes as "entities that are psychologically autonomous and that constitute institutional persons" (Pettit, 2010, 167).

I argue however that groups do not have minds of their own: rather they are social constructions/tools that are ultimately dependent upon the at least tacit support of individuals. We may regard them as distinct minds or entities for practical or legal purposes but in all cases they are social constructions dependent on their members.

Carole Rovane (1997) proposes a normative account of personal identity which supports ethical equality between group persons and individual persons. On this view, what makes someone a person is his or her striving to achieve overall rationality in his or her beliefs and actions. Rational unity is

never a metaphysical given but always a product of effort and will. This rational unity can be achieved within parts of human beings, within whole human lives, but also sometimes at the level of a group. On Rovane's view then, groups acting under a specific commitment to rationality are persons in their own right.

Rovane (2014) recognizes that the rational unity that is achieved at the level of the group typically precludes rational unity at the level of its human constituents within their whole lives, though it can be realized within parts of those human lives. In spite of this, she rejects Pettit and List's metaphysical and normative individualism, on which view, human beings qualify as *natural persons*. On Rovane's view all persons are to be treated equally – whether they are group persons or individual persons – because the sort of rational unity that defines individual rational unity is realized at the level of a whole group.

However Pettit agues, rightly I think, that natural persons have "an inescapable priority' because it is always up to them "whether or not to cede place to the institution, acting in furtherance of a collective goal in neglect of his or her own priorities" (Pettit, 2010, 190). On this view then the ethical priority remains with the individuals.

Max Scheler also supported the idea that there are collective persons (Salice, 2016). According to Scheler there are three forms of sociality (emotional contagion, co-experiencing and empathy) and three corresponding kinds of group (masses, communities and societies). Collective persons are special kinds of communities. According to Scheler, the default situation for communities is not one in which there is an isolated *I* who joins others to form a *we.* Rather communities start from a situation in which the *we* encompasses the *I* so pervasively that the *I* is fully identified with the *we.* Communities can become collective persons when their members come to an understanding of themselves as being *one* individual and *one* specific member of the community among other individuals but also specific members of the community. So collective persons are communities whose members conceive of themselves as individual persons and who are also specific members of their community. For Scheler then, personhood, for

individuals, is not conferred by birth; rather it is a status that can be acquired by maturity. And, by becoming fully aware of being owners of their own mental states, and also co-owners of collective mental state, members of a community can change the nature of their community into a collective person. What most distinctively characterizes the collective person is its sense of solidarity. Each member of the community is not only fully responsible for his or her actions, but is also co-responsible for the actions of others and of the community.

On this account however it seems that the 'collective person' is better understood as a conscious construct of individuals that needs to be bolstered by empathy and solidarity. While his account of the historical or developmental priority of the community may be correct, it would seem that once individuals become fully aware of being owners of their own mental state it would be unlikely that they would cede sufficient authority to the community such that it would have a status of personhood equal to their own. There will always be limits to empathy and solidarity.

Could any of these accounts of collective minds or collective persons support the existence of collective selves over and above individuals? Zahavi and Salice (2016, 525) think not. While there are multiple types of *we,* or as Scheler puts it, forms of being-together, they are all based on a fundamental form of dyadic interaction. As Zahavi puts it –

> "prototypical forms of we-intentionality are founded upon empathy and joint attention and require a special reciprocal second-person perspective. The self-other differentiation, the distinction between self and other, consequently precedes the emergence of, and is retained in the we" (Zahavi, 2014, 248).

On this view then a foundational component of all social entities is a basic empathic capacity to directly engage in dyadic relationships with other minded creatures. The you-me relation is always a necessary component of any we.

Also, having a mind or being a person (at least as defined by Scheler or Rovane) does not exhaust the self that we know by being it. None of the proposed collective entities are conscious, in the way that we are. As Pettit

put it – if groups are persons then they are persons of " a bloodless, bounded, and crudely robotic variety" (Pettit, 1993, 188) or "zombies: minds with no consciousness" (O'Madagain, 2012, 282). A collective mind or person therefore does not amount to a self – in the sense of a self that we know ourselves to be.

6.5 ELEPHANTS, OCTOPUSES, CELLS & SELVES

I started this chapter by using the foil of counting how many selves there are to examine some theories that suggest there are none. I argued that the kind of selves these theories are rejecting do not coincide with, or are not inconsistent with, the kind of self that I am talking about. Also, I argue that thinking that conscious selves are an illusion is self-contradictory; while the theory that there is just one self is not consistent with the kind of self we know ourselves to be. So on my count, accepting the existence of other human beings like me, there are at least 7.8 billion selves. In the last section I considered, and rejected, the idea that two or more people could form a group in such away that the group itself would constitute a self in the sense in which we are selves.

In this chapter so far therefore, I have being using the concept of a self like us humans as the unit of counting; I'll use self(h) or selves(h) when I mean a self in this sense. Earlier in the dissertation, especially in chapter 4 on biology, I have used selves in a wider way to include all biological individuals from cells to organisms; which have certain self-individuating structure; I'll use self(b) or selves(b) for this sense.

In this dissertation I argue that self(h) is the primary sense of self in that its what we know by being it; and that we use our knowing of self(h) to identify other entities – selves(b) with the minimal features that could count as a self. These two senses of self can be seen as the ends of a spectrum and we can identify different intermediate selves, as I'll explain in this section. So there are at least 7.8 billion selves(h): and dramatically many more

selves(b) - in one human body it is estimated that there are 37.2 trillion cells (Bianconi, 2013) – and then there are all the other living things made up of cells. If you could do the maths, the count for selves(b) is going to be a very big, but not infinite, number.

The selves(h) count could be increased if it could be shown that there are other creatures sufficiently like us in some relevant ways so that we would have to count them as selves like us. The first place to look for possible candidates is to our nearest living genetic relatives - the closest species to humans on the tree of life. Thus primates such as chimps, bonobos and gorillas seem like obvious candidates to share aspects of our selfhood. One primary fact about ourselves is that we are persons with rational capacity. Don Ross (2018, 2019) argues that to be a person in this sense involves having an ability to communicate with language and having the ability to regulate your spontaneous emotional reactions under the influence of explicit social norms. Although chimps, bonobos and gorillas cannot talk because they lack suitably manipulable vocal tracts, some individuals have been taught to use keyboards or sign language. However, no ape has managed to exceed the language skills of very young humans, or has ever used language to do anything other than express a relatively immediate *desire*. From this some theorists argue that if our nearest living relatives aren't naturally prepared for language, then this is a good basis for conjecturing that *no* nonhumans are. This suggests that only humans can be persons – so the current selves(h) count should remain at 7.8 billion.

However Don Ross (2018, 2019) suggests that this analysis is wrong because it ignores *convergent evolution*: the fact that natural selection often stumbles across the same capacity independently on different branches of the tree of life.

As he points out –

> "The closest living relatives of birds are lizards; lizards cannot fly; but that doesn't license an inference that *no* nonbirds can fly. Likewise, the fact that our closest living relatives don't use language does not justify the conclusion that no other creatures can."

Ross argues that the most important dimension on which hominins split off from other apes was in adopting *hypersociality*: that is, in developing brains that are specially adapted to coordinating thoughts and actions with others. This hypersociality coevolved with language. Although apes are social, they are not hypersocial and therefore, it is not surprising that they don't have brains that are prepared for language.

In searching for species that might demonstrate non-human language use, Ross suggests two criteria –

"First, we need behavioural evidence for flexible, intelligent hypersociality that *could* make language development a viable investment for natural selection. Second, we need evidence of information exchange, using signals that have enough acoustic variation to make it physically possible that they could encode syntax." He points out that dogs meet the first criteria but not the second and many songbirds might meet the second but not the first. He argues that there is a loose consensus among comparative psychologists that the zone of possibility for both criteria currently boils down to the following animals: parrots, corvids (crows, ravens, jays), toothed whales (dolphins, porpoises, sperm whales, orcas) – and elephants.

Peter Godfrey-Smith (2016) also points to convergent evolution to argue that our search for other minds should not be confined to genetically close species. Octopuses, cuttlefish and squid are very distant from us on the tree of life: our common ancestors were probably "small flattened worms" that lived in the ocean about 600 million years ago (in comparison our common ancestor with elephants lived about 160 million years ago). Yet Godfrey-Smith claims that he has had "a sense of mutual *engagement*" with them (Godfrey-Smith, 2016, 5).

Godfrey-Smith contention is that complex brains evolved separately at least three times – once in animals like us but also separately in octopuses and squids. He claims that recent evidence from genetics shows that the evolutionary split between the ancestors of octopuses and squids occurred about 270 million years ago and that they developed in parallel after that (Ibid.196). So although octopuses and squids appear in some ways very

different or alien to us, their interaction with us suggests that they have minds and we would want to count them as selves.

One way that I have characterized the kind of self we are is as conscious subjects of experience. Godfrey-Smith identifies subjective experience as the most basic phenomenon common to all animals. On this view, our consciousness is one form of subjective experience, but it is not the only form. He argues that there is something about the way animals deal with the world that makes a crucial difference and brings about a subject of experience or self. In animal evolution, along with the sheer elaboration of sensing and acting, there's the evolution of new kinds of connection between these activities, especially connections that loop, and therefore involve feedback. He points out that loops that link actions back to the senses are not only seen in us. They are present also in very simple forms of life. But they become more marked in animals, especially because animals can do more –

> " The evolution of muscle, derived from tiny fiber-like elements inside cells, created a new means by which life impresses itself on the world… It makes possible the *manipulation* of objects, the deliberate and rapid transformation of what is around us" (Godfrey-Smith, 2016, 82).

The evolution of animals is affected by these looping causal paths in a number of ways: the animal must distinguish between 'self' and 'other' – between the effects on its senses of its own actions and the effects of events going on around it; the interaction between perception and action is also seen in *perceptual constancies* – an object can remain recognizable as the same object while our viewpoint on it changes; and as streams of information come in from different senses they are integrated into a single picture (Ibid., 83-84). Thus, Godfrey-Smith argues that all animals are to some degree inevitably unified selves: as living agents, animals are physical objects keeping themselves alive (Ibid. 87).

Don Ross's and Peter Godfrey-Smith's identification of creatures who are like us in certain ways suggest that these creatures are selves that are close to, but maybe not coincident with selves(h) on the spectrum of selves. Also, the spectrum is not singular – elephants may be close to us in terms of

sociality and language, while octopuses and squids may be close to us in terms of intelligence – while we share the capacity to have subjective experiences with all animals.

The evidence from biology suggests that we can extend the spectrum of selves not just to all animals, but also to all living entities. As I discussed in Chapter 4 Thomas Pradeu (2012) points to the significance of the immune system in delineating biological individuals or organisms. He defines an organism as "a functionally integrated whole, made up of heterogeneous constituents that are locally interconnected by strong biochemical interactions and controlled by systemic immune interactions that repeat constantly at the same medium intensity" (Pradeu, 2010, 16). Traditionally, it was thought that only jawed vertebrates possessed a true immune system. Thus F.M. Burnet (1961) thought that only the immunity provided by lymphocytes was effectively specific, and thus constituted a genuine immunity and secondly, in a vision of immunology that was marked by the significance of vaccination phenomena, he considered that actual immunity existed only where there was immune memory, and this only evolved with jawed vertebrates. However, Pradeu points out that most immunologists today consider, contrary to Burnet, that it is possible to speak of immune systems in all animals, vertebrates and invertebrates alike (Pradeu, 2012, 23). This is the result of the identification of a wider range of immune mechanisms including immune memory mechanisms. And Pradeu argues that immune systems can also be found in plants and further that immune systems exist in all organisms, both unicellular and multicellular –

> " …everywhere in the living world, and opposed to what was believed just ten or fifteen years ago, immune mechanisms are at work. In one sense, that should not be surprising, since all organisms, even the simplest, are subjected to pathogen actions" (Pradeu, 2012, 25).

If, as Pradeu argues, "immunity is a ubiquitous phenomenon in nature" (Pradeu, 2012, 25), then, biological individuals, what I consider the basic form of self – self(b), are also ubiquitous in living nature. This is consistent with Hans Jonas (1966) and Evan Thompson's (2007) argument that any piece of the world that responds to stimuli and controls the flow of

molecules across its borders, is a self in a basic sense. On their view then even cells, which are the basic building blocks of all life, have a core set of organizational properties that justifies identifying them as selves(b). Thompson argues that what Immanuel Kant (1790) recognized as a distinguishing characteristic of organic beings – that they are unities rather than mere aggregates – finds minimal expression in a living cell –

> "A cell, not merely a persisting material aggregate, is a self-sustaining unity, a unity that dynamically produces and maintains its own identity in the face of what is other" (Thompson, 2007, 149).

This is consistent with Hans Jonas's identification of a self with internal identity, as the distinctive mark of life –

> "The introduction of the term 'self', unavoidable in any description of the most elementary instances of life, indicates the emergence, with life as such, of internal identity – and so, as one with that emergence, its self-isolation too from all the rest of reality" (Jonas, 1966, 82-83).

What connects selves(h) and selves(b) on the one spectrum is that we, from our selves(h) perspective, can identify many of our significant characteristics as being prefigured in the organizational structure of selves(b). Hans Jonas takes the rather surprising examples of freedom and teleology, as a human characteristic that are evident in selves(b). As he explains –

> "One expects to encounter the term ['freedom'] in the areas of mind and will, and not before: but if mind is prefigured in the organic from the beginning, then freedom is" (Jonas, 1966, 3).

Jonas follows Immanuel Kant (1790) in taking the organism's self-organizing way of being as his starting point. However for Kant, although organisms are natural purposes because they are self-organizing beings, they are not free in that they self-organize according to a predetermined pattern, that we cannot know. Therefore, we human beings, in so far as we are considered as products of nature, are not free: it is only as rational moral agents who transcend nature that we are free. For Kant, freedom belongs to the sphere of practical reason in which reason must presuppose its own

freedom in order for morality to be possible. Thus, for Kant, we transcend nature by means of our rational being and not in virtue of our organic being.

For Jonas, on the other hand, it is the individuality or selfhood of the organism that links freedom as an "objectively discernible mode of being" in the living world and freedom in the sense or human liberty. According to Jonas the term *freedom* describes "a certain independence of form with respect to its own matter" (Jones, 1966, 81) achieved through metabolism. As Steven Rose (1997) argues, without metabolism, there can be no "lifeline" – no developmental continuity of the individual through material change. According to this view then, every organism enacts a lifeline and thus achieves a kind of freedom in relation to the materiality of the world. Thus, an organism's identity is accomplished dynamically at a formal level and is not bound to its material constitution because this constitution is constantly renewed. However, this freedom comes with a correlative necessity; the organism has to constantly change. Jonas therefore describes the relationship of the organic form to matter as *needful freedom.*

Furthermore, for Jonas metabolism establishes a self with an internal identity, and operates according to internal norms that determine whether otherwise neutral events are good or bad for the continuation of the organism. In these ways he argues that metabolism is immanently teleological. As he puts it –

> "Organic individuality is achieved in the face of otherness, as its own ever challenged goal, and is thus teleological" (Jonas, 1968, 243).

For Jonas then, the kind freedom and teleology discernible at level of the organism is linked to freedom and teleology at the human level by the spectrum of selfhood. It is because we know ourselves as selves with freedom and purpose as primary facts of our existence, that we can discern freedom and purpose at the level of organism. And Jonas points to the significance of the evolution of selves(b) in the history of the universe –

> "… it is in the dark stirrings of primeval organic substance that a principle of freedom shines forth for the first time within the vast necessity of the physical universe – a principle foreign to suns, planets, and atoms" (Jonas, 1966, 3).

The significance of the evolution of selves(b) for the history of the universe, and the significance of selves(h) for the future of the universe, is the topic of the next chapter – 'Why selves count'.

CHAPTER 7. WHY SELVES COUNT

7.1 Anthropic Analysis
7.2 The Weak Anthropic Principle
7.3 Explanatory Knowledge and Judgment

In the last chapter I rejected the ideas that selves are an illusion, that there are none or only one self, or that groups could be selves in their own right. More positively, I endorsed the idea that selves(b) are co-extensive with life. This suggests that the appearance of selves(b) was the most important event in the history of the universe, and points to the significance of humans as conscious selves(h) for the future of the universe. In this chapter I argue that it is our human capacity for judgment that enables us to achieve genuine explanatory knowledge. This shows our cosmic significance as a source of nature's self-understanding – why selves count.

7.1 ANTHROPIC ANALYSIS

The anthropic principle expresses the idea that it is not surprising that the observed Universe admits the existence of life, for the Universe cannot be observed unless there are observers in it.

An early version of this idea can be found in the pre-Socratic philosopher Protagoras. Only a few brief quotations from Protagoras have survived, embedded in the works of later authors. Perhaps the most famous is the *homo-mensura* (man-measure) statement –

"Of all things the measure is man, of the things that are, that [or "how"] they are, and of things that are not, that [or "how"] they are not." (DK80b1)

As with any short extract taken out of context, the precise meaning of this statement is not necessarily obvious. While later thinkers have interpreted this in different ways, one central idea seems to point to the significance of man in understanding the world.

David Deutsch (2011) points out that Pre-Enlightenment thinking was anthropocentric in two mutually supporting ways. Firstly, humans are placed at the physical hub of the universe. Secondly, explanations were anthropocentric; explanations were in terms of "entities with intentions and human like thoughts, including supernatural people such as spirits and gods. So, winter might be attributed to someone's sadness, harvests to someone's generosity, natural disasters to someone's anger, and so on" (Deutsch, 2011, 42). Because these explanations often involved cosmically significant beings caring about what humans did, or having intentions about them, this conferred cosmic significance on humans too, if only derivatively.

In the sixteenth century, however, Copernicus dethroned humans from their central place in the scheme of things, and he thereby came to be seen as the initiator of the Scientific Revolution. As John A. Wheeler puts it, Copernicus' "model of the motions of the planets and the Earth taught us to look at the world as machine. Out of that beginning has grown a science which at first sight seems to have no special platform for man, mind, or meaning" (Foreword to Barrow and Tipler, 1986, vii).

We now know that we are not at the centre of the universe; in fact, the universe appears to be homogeneous in all directions so that it does not have a geometric centre. Although some of the astrophysical phenomena played a significant role in our past, we now know that we have never been significant to them. So it may seem that human beings and their wishes and actions are extremely insignificant in the universe at large. The physicist, Carlo Rovelli writes that he likes physics precisely because it is not anthropocentric – "We are obsessed with ourselves. We study *our* history, *our* psychology, *our* philosophy, *our* gods. Much of our knowledge revolves around man himself, as if we were the most important thing in the universe." Physics, on the other hand, "opens a window through which we can see further" and "gives … the sense of fresh air entering the house" (Rovelli, 2017, xi).

David Deutsch (2011) argues that scientific anti-anthropocentricism generated a new conceptual framework that informed an entire world view, based on two ideas –

- the Principle of Mediocrity: there is nothing significant about humans (in the cosmic scheme of things). As the physicist Stephen Hawking (1995) put it, humans are "just a chemical scum on the surface of a typical planet that's in orbit round a typical star on the outskirts of a typical galaxy".

- Spaceship Earth – a metaphor for the biosphere: outside the spaceship (biosphere), the universe is implacably hostile, but the interior is a vastly complex life-support system. The biosphere has the 'appearance of design'. It seems to be highly adapted to sustaining us, but this is only because we were adapted to it by evolution. However, its capacity is finite; if our lifestyle exceeds its capacity we will ruin our life support system and we will have nowhere else to go.

Deutsch claims that both these ideas suggest "that the reach of the distinctively human way of being – that is to say, the way of problem solving, knowledge-creating and adapting the world around us – is bounded" (Deutsch, 2011, 54). And, both these ideas suggest that the bounds of our reach cannot be very far beyond what has already been reached, so that any attempt to go beyond that range will likely lead to failure and catastrophe respectively.

This perspective prioritizes a scientific view of the world over any human dimension. As Wilfrid Sellars put it, referencing Protagoras -

> "Speaking as a philosopher, I am quite prepared to say that the common-sense world of physical objects in Space and Time is un-real – that is, that there are no such things. Or, to put it less paradoxically, that in the dimension of describing and explaining the world, science is the measure of all things, of what is that it is, and of what is not that it is not" (Sellars, 1997, 83).

Questioning of this perspective started in 1957, with Princeton physicist Robert Dicke (1957), who argued that for advanced carbon-based life to

exist, the Universe had to be roughly the age that it is now. If it were much younger, and there would not have been time for sufficient interstellar levels of carbon to build up by nucleosynthesis; if it were much older, and the golden age of main sequence stars and stable planetary systems would have drawn to a close.

Brandon Carter (1974) named this idea the anthropic principle, and specifically understood it as a "reaction against exaggerated subservience to the 'Copernican principle'".

> "Copernicus taught us the very sound lesson that we must not assume gratuitously that we occupy a privileged *central* position in the Universe. Unfortunately there has been a strong (not always subconscious) tendency to extend this to the most questionable dogma to the effect that our position cannot be *privileged* in any sense" (Carter, 1974, 291).

Carter argues that this dogma is clearly untenable if one accepts (a) that specially favourable conditions (of temperature, chemical environment, etc.) are prerequisite for our existence, and (b) that the Universe evolves, and is by no means spatially homogeneous on a local scale. Thus, Carter proposes "the *anthropic principle* to the effect that what we can expect to observe must be restricted by the conditions necessary for our presence as observers." So, although our position is not necessarily *central,* it is inevitably *privileged* to some extent.

Carter identified two versions of this principle –

- "the '*weak' anthropic principle* - we must be prepared to take account of the fact that our location in the universe is *necessarily* privileged to the extent of being compatible with our existence as observers" (Ibid. 293).
- "the '*strong' anthropic principle* - the Universe (and hence the fundamental parameters on which it depends) must be such as to admit the creation of observers within it at some stage. To paraphrase Descartes – *Cogito ergo mundus talis est"* (Ibid. 294).

Carter uses these principles to argue that certain "large number coincidences" in physics and cosmology can in principle be predicted in

advance of their observation. He uses the weak anthropic principle to predict that the Hubble fractional expansion rate of the universe is equal to, within a few powers of ten, the reciprocal of the same large number (Ibid. 292). He uses the strong anthropic principle to show a priori that if the universe is closed, its present age *t* might already be comparable with its total lifetime T (Ibid. 293).

John Barrow and Frank Tipler (1986) proposed a more extensive taxonomy of anthropic principles –

- Weak Anthropic Principle (WAP): The observed values of all physical and cosmological quantities are not equally probable but take on values restricted by the requirement that there exist sites where carbon-based life can evolve and by the requirement that the Universe be old enough for it to have already done so.
- Strong Anthropic Principle (SAP): The Universe must have those properties which allow life to develop within it at some stage in its history. (Ibid. 21)

SAP can be interpreted in different ways –

- *(A) There exists one possible Universe 'designed' with the goal of generating and sustaining 'observers'.*

- *(B) Observers are necessary to bring the Universe into being. (Wheeler's Participatory Anthropic Principle, PAP).*

- (C) An ensemble of other different universes is necessary for the existence of our Universe (Ibid. 22).

Barrow & Tipler propose the following generalization of the SAP –

- Final Anthropic Principle (FAP): Intelligent information processing must come into existence in the Universe, and once it comes into existence, it will never die out (Ibid. 23). Barrow & Tipler say that the FAP implies a melioristic evolving cosmos (ibid. 675), and their book ends with the following prediction -

  "At the instant the Omega Point is reached, life will have gained control of *all* matter and forces not only in a single universe, but in all universes

whose existence is logically possible; life will have spread into *all* spatial regions in all universes which could logically exist, and will have stored an infinite amount of information, including *all* bits of knowledge which it is logically possible to know. And this is the end" (ibid. 677).

Reviewing this increasingly speculative array of principles, Martin Gardner (1986) satirically proposed that the last principle is best called the Completely Ridiculous Anthropic Principle (CRAP for short!).

In fairness, Barrow and Tipler themselves warned that "both the FAP and the SAP are quite speculative; unquestionably, neither should be regarded as well-established principles of physics. In contrast, the WAP is just a restatement, albeit a subtle restatement, of one of the most important and well-established principles of science: that it is essential to take into account the limitation's of one's measuring apparatus when interpreting one's observations" (Ibid. 23). This may be an unduly dismissive assessment if it means to equate the 'observer' with 'measuring equipment' – a person who is able to take into account the limitations of measuring equipment is not just another piece of measuring equipment. In any case, it is the WAP that has been most productive in stimulating further work.

7.2 THE WEAK ANTHROPIC PRINCIPLE

According to Barrow and Tipler, the Weak Anthropic Principle states that any cosmological observations made by astronomers are biased by an all-embracing selection effect: our own existence. Therefore, "features of the Universe which appear to us astonishingly improbable, a priori, can only be judged in their correct perspective when due allowance has been made for the fact that certain properties of the Universe are necessary if it is to contain carbonaceous astronomers like ourselves" (Barrow and Tipler, 1986, 15). They stress that this statement "is in no way either speculative or controversial. It expresses only the fact that those properties of the Universe

we are able to discern are self-selected by the fact that they must be consistent with our own evolution and present existence" (ibid. 16).

It seems that a great deal of "fine tuning" is required in order that life be possible. In particular various fundamental constants are highly constrained in their values if life as we know it is to exist. As M.J. Rees (2003) explains -

> "A universe hospitable to life – what we might call a biophilic universe – has to be special in many ways … Many recipes would lead to stillborn universes with no atoms, no chemistry, and no plants; or to universes too short lived or too empty to evolve beyond sterile uniformity."

So it seems that only particular laws of physics, and particular initial conditions in the Universe, allow for the existence of intelligent life, as we know it. Paul Davies (2006) called this the Goldilocks Enigma – why is the universe just right for life?

The Weak Anthropic Principle can be seen as the somewhat self-referential answer to this question - it is not surprising that the observed Universe admits the existence of life, for the Universe cannot be observed unless there are observers in it. And this self-referential answer allows us to invert the question. As David P. Barash (2018) puts it –

> "… maybe it's not surprising that we live in a Universe suitable for life, not because that Universe has been fine-tuned for us … but because we are fine-attuned to it as a result of natural selection. Just as the physical qualities of air have selected for the structure of bird wings, and the anatomy of fish speaks eloquently about the nature of water, maybe the nature of the physical Universe has in the most general sense, selected for life, and thus, for us."

Thus, George Ellis (2008) observes that this seemingly empty statement gains content when we turn it round and ask, at what times and places in the Universe can life, as we know it, exist; and what are the inter-connections that are critical for its existence.

> "It could not for example exist too early in the present expansion phase, for the night sky would then have been too hot. Furthermore one can

> deduce various necessary relations between fundamental quantities in order that the observers should exist, so that if for example the fundamental constants vary with time or place in the Universe, life will only be possible in restricted regions where they take appropriate Anthropic values" (Ellis, 2008, 9.1.4).

Hence, Ellis says, this view basically interprets the Anthropic principle as a selection principle; the necessary conditions for observers to exist restricts the times and places from which the Universe can be observed.

The weak anthropic principle can thus be seen as a constraint on physics and physicalism, and therefore on any purely scientific metaphysics. According to Hellman and Thompson (1975, 551) –

> "Mathematical physics, as the most basic and comprehensive of the sciences, occupies a special position with respect to the over-all scientific framework. In the loosest sense physicalism is a recognition of this special position."

As I mentioned in chapter 3, Ladyman and Ross formulate this asymmetry into a rule – the "Primacy of Physics Constraint", which they state as -

> "Special science hypotheses that conflict with fundamental physics, or such consensus as there is in fundamental physics, should be rejected for that reason alone. Fundamental physical hypotheses are not symmetrically hostage to the conclusions of the special sciences" (Ladyman and Ross, 2007, 44).

It might be argued that the anthropic principle is not a principle of a special science and therefore there is no conflict with the Primacy of Physics Constraint. However, the anthropic principle is a constraint on any possible metaphysical theory. And any metaphysical theory must allow for the existence of a metaphysician (a person with the ability to produce a metaphysical theory).

As I mentioned in chapter 5, Ladyman and Ross understand metaphysics as the enterprise of producing a relatively unified picture of the world by

critically elucidating consilience networks across the sciences. With this in mind they propose two verificationist claims -

- No hypothesis that the approximately consensual current scientific picture declares to be beyond our capacity to investigate should be taken seriously.
- Any metaphysical hypothesis that is to be taken seriously should have some identifiable bearing on the relationship between at least two relatively *specific* hypotheses that are either regarded as confirmed by institutionally *bona fide* current science or are regarded as motivated and in principle confirmable by such science. (Ladyman and Ross, 2007, 29)

If the anthropic principle is taken into account then we might need another claim - No metaphysical hypothesis should be taken seriously that disallows the existence of metaphysicians who can critically elucidate consilience networks across the sciences.

Another way to turn around the Weak Anthropic Principle is to ask what properties an observer must have if the Universe is successfully observed. Ellis approaches this by asking whether the ultimate nature of existence is purely material, or does it embody some form of rationality ('Logos') and/or purpose ('Telos'). Ellis argues that even in order to understand just the material world, it is necessary that one needs to consider forms of existence other than the material only – for example a Platonic world of mathematics and a mental world, both of which can be claimed to exist and be causally effective in terms of affecting the material world. Thus our understanding of local causation will be incomplete unless we take them into account.

Furthermore Ellis argues that in examining these issues one needs to take into account data about the nature of our existence that comes from our daily lives and the broad historical experience of humanity (our experience of ethics and aesthetics, for example), as well as those discoveries attained by the scientific method. Part of this data, Ellis argues, is the evidence of our actions in trying to understand the world -

> "Many writings claim there is no purpose in the universe: it is all just a conglomerate of particles proceeding at a fundamental level in a purposeless and meaningless algorithmic way. But I would reply, the very fact that those writers engage in such discourse undermines their own contention; they ignore the evidence provided by their own actions. There is certainly meaning in the universe to this degree: *the fact they take the trouble to write such contentions is proof that they consider it meaningful to argue about such issues*; and this quality of existence has emerged out of the nature of the physical universe" (Ellis, 2008, 9.3.4).

Thus Ellis argues that the human mind is causally effective in the real physical world precisely through many activities motivated by meanings perceived by the human mind. Therefore, any attempt to relate physics and cosmology to ultimate issues must take such real world experience seriously, otherwise it will simply be ignoring a large body of undeniable data about dimensions of existence that indeed do occur.

7.3 EXPLANATORY KNOWLEDGE AND JUDGMENT

This interpretation of the anthropic principle suggests that our actions as observers and explainers of our world exemplify a form of existence that is significant, not just for ourselves, but also for the universe as a whole. David Deutsch (2011, 45) develops this idea with two claims that challenge the scientific view of the world –

- People *are* significant in the cosmic scheme of things (contradicting the Principle of Mediocrity); and

- The Earth's biosphere is *incapable* of supporting human life (contradicting the Spaceship-Earth metaphor).

Contrary to Hawking (1995), Deutsch argues that although we are on a (somewhat) typical planet of a typical star in a typical galaxy, we are far from typical of the matter in the universe. What makes humans significant is our ability to gain explanatory knowledge. He argues that "The ability to create and use explanatory knowledge gives *people* a power to transform nature which is ultimately not limited by parochial factors, as all other adaptations are, but only by universal laws. This is the cosmic significance of explanatory knowledge – and hence of people, whom I shall henceforward define as entities that can create explanatory knowledge." (Deutsch, 2011, 56) Thus he argues that today, contrary to the spaceship earth metaphor of the biosphere, almost the entire capacity of Earth's 'life support system for humans' has been provided not *for* us but *by* us, using our ability to create new knowledge.

Brian Cantwell Smith (2019) clarifies the nature and significance of human explanatory knowledge; he claims it involves a capacity for judgment, which I identified in Chapter 1 as one of the primary facts of the way we find ourselves living. Smith disagrees with Thomas Hobbes's claim that

"…'reason' in this sense [as a faculty of the mind] is nothing but 'reckoning,' that is adding and subtracting, of the consequences of general names agreed upon for the 'marking' and 'signifying' of our thoughts" (Hobbes, 1651, Ch. V).

Smith uses 'reckoning', as the term for the calculative rationality that present-day computers and AI systems use. And human judgment is also more than the registrational powers of animals, which do not extend beyond the ways in which they conduct their lives, what they care about, and what they are vulnerable to. Over and above this, Smith characterises human judgment as "a form of dispassionate deliberative thought, grounded in ethical commitment and responsible action, appropriate to the situation in which it is deployed" (Smith, 2019, xv). His claim is that this is the normative ideal to which adult humans have the capacity to aspire – not that every human cognitive act meets this ideal.

Deutsch endorses Karl Popper's (1959) rejection of accounts of science in terms of induction and empiricism. We do not passively derive our knowledge from the content of our senses. Scientific theories are not 'derived' from anything - we do not read them in nature, nor does nature write them into us. Rather, they are guesses – bold conjectures (Popper, 1963). Human minds create them by rearranging, combining, altering, and adding to existing ideas with the intention of improving upon them.

Popper's characterization of scientific theories as human conjectures is similar in some ways with the more recent idea of predictive coding developed by Friston (2010), Hohwy (2013) and Andy Clark (2013, 2016). This theory depicts human brains as devices that minimize prediction error signals: signals that encode the difference between actual and expected sensory stimulations.

But as Andy Clark (2017) now argues - "We humans often seem to actively seek out surprising events, deliberately harvesting novel and exciting streams of sensory stimulation". This suggests that "we must go beyond the discussion of simple information-theoretic imperatives (such as 'minimize long-term prediction error') and … recognize the essential role of species-specific prestructuring, epistemic foraging, and cultural practices in shaping the restless, curious, novelty-seeking human mind".

Deutsch argues that experience is essential to science, but its role is different from that supposed by empiricism. It is not the source from which theories are derived. Rather, experience is used to choose between theories that have already been guessed. Popper argues that falsifications are particularly important because "[t]hey teach us the unexpected; and they reassure us that, although our theories are made by ourselves, although they are our inventions, they are none the less genuine assertions about the world; for they clash with something we never made" (Popper, 1972, p. 197). As Piaget (1980) points out we learn about reality by experiencing the obduracy of the world. That is what 'learning from experience' is.

This is consistent with Brian Cantwell Smith's claim that human judgments are accountable to the world. As he explains it, our registrations of the world are necessary abstractions from the world's detail: because of

our finite brain capacity and our partial disconnection from the world we need representations to be able to orient to anything beyond our effective grasp. But he argues – "… no matter their necessity, our *registrations* are not what matters – what matters is *that which we register.* To be accountable, to hold things to account, is to know the difference between the two – and to be committed to the later, not the former" (Smith, 2019, 112).

Deutsch does not limit the domain of explanation just to science - "The quest for good explanation is, I believe, the basic regulating principle not only of science, but of the Enlightenment generally. … it … implies a methodological rule – *a criterion for reality* – namely that we should conclude that a particular thing is real if and only if it figures in our best explanation of something" (Deutsch, 2011, 22-23).

Deutsch points out that using knowledge to cause automated physical transformations is, in itself, not unique to humans. This is the basic method that all organisms use to keep themselves alive; every cell is a chemical factory. The difference between humans and other species is in the kind of knowledge they can use (explanatory instead of rule-of-thumb) and how they create it (conjecture and criticism of ideas, rather than the variation and selection of genes). It is precisely those two differences that explain why every other organism can function only in a certain range of environments that are hospitable to it, while humans transform *in*hospitable environments like the biosphere into support systems for themselves. Thus, he argues that "while every other organism is a factory for converting resources of a fixed type into more such organisms, human bodies (including their brains) are factories for transforming *anything into anything* that the laws of nature allow. They are 'universal constructors'" (Deutsch, 2011, 58-59).

Deutsch argues that from the beginning, it was human knowledge that made the planet even marginally habitable by humans, and the enormously increased capacity of our life-support since then (in terms both of numbers and of security and quality of life) has been entirely due to the creation of human knowledge. As Ted Nordhaus (2018) points out –" we have been engineering our environments to more productively serve human needs for tens of millennia". This means that the carrying capacity of our planet is not

fixed and there is no particular reason to think that we won't be able to continue to raise carrying capacity further. As Deutsch points out, since the Enlightenment, technological progress has depended specifically on the creation of explanatory knowledge. … "This increasingly intimate connection between *explaining* the world and *controlling* it is no accident, but is part of the deep structure of the world" (Deutsch, 2011, 55).

According to Deutsch our ability to explain and control makes humans the most significant cosmic entities -

> "From the least parochial perspectives available to us, *people* are the most significant entities in the cosmic scheme of things. They are not 'supported' by their environments, but support themselves by creating knowledge. Once they have suitable knowledge, they are capable of sparking unlimited further progress" (Deutsch, 2011, 76).

Although Smith argues that there are profound differences between the kind of judgment that humans are capable of and the sorts of reckoning systems we at present have the ability to construct, he does not believe that artificial general intelligence is inherently beyond the reach of synthetic construction. If we are ever capable of doing this, it will be a human achievement; but it will not be in the short term. In the mean time he says "… we should stand in awe of the capacity of the human mind, and of the achievements of human culture, in having developed registrational strategies, governing norms, ontological commitments, and epistemic practices that allow us to comprehend and go to bat for the world as world" (Smith, 2019, 147). Our knowing-by-being of ourselves as the source of this knowledge allows for the possibility that we might use this knowledge wisely.

Starting from the anthropic fact of our own existence we can infer certain aspects of our universe; it must be such that it allows for our existence, with dimensions other than the physical. And from the way we find ourselves acting in, and explaining the world, we know certain aspects of our own existence, and therefore science is constrained. As George Ellis points out above, part of the way we come to know ourselves is from the very action

we are involved in when we try to understand our place in the world. And as Deutsch and Smith argue our capacity for judgment enables us to create genuine explanatory knowledge, and this gives us cosmic significance such that we can realize our responsibility as a source of nature's self-understanding.

CHAPTER 8. THE PRIMACY OF BEING ME

8.1 Individual Conscious Subjects
8.1.1 The Universe and Me
8.1.2 Being a Conscious Subject
8.1.3 Being Me
8.2 Objectivity, Existential Primacy and Ontology
8.2.1 Objectivity
8.2.2 Existential Primacy
8.2.2 The Ontological Status of Consciousness
8.3 The view from here.

8.1 INDIVIDUAL CONSCIOUS SUBJECTS

8.1.1 THE UNIVERSE AND ME

The anthropic analysis of the last chapter can lead to a tension in understanding our place in a wider world. As Thomas Nagel put it in *The View from Nowhere,* the problem is –

"how to combine the perspective of a particular person inside the world with an objective view of that same world, the person and his viewpoint included" (Nagel, 1986, 3).

Although, all our knowledge is from a particular point of view, nevertheless we seem to be able to conceive of the world from an objective standpoint that seems to the deny the significance of any particular perspective.

In this final chapter I return to the significance of our first-person experience of ourselves as individual conscious subjects, and the importance of this for a proper understanding of our place in nature.

8.1.2 BEING A CONSCIOUS SUBJECT

The analytic philosophical discussion about consciousness started by focusing on the subjective feel or 'what it is like' to have a particular experience. Thomas Nagel's early (1974) paper was titled "What is it like to be a bat?" More recently some philosophers have tried to shift the focus of attention to another feature of conscious experiences – the fact that 'what it is like' to have a conscious experience always involves *someone.* It is argued that a fuller description of a conscious experience is that 'there is something it is like for someone' to undergo any particular experience. Joseph Levine (2001, 7) calls this the "subjectivity" of experience. Uriah Kriegel (2009) uses the terms "for-me-ness" or "subjective character". And, as I pointed out in chapter 2, the phenomenologists have always argued that a pre-reflective self-awareness accompanies all our conscious experiences, and thus the subject is never separated from any experience.

But what does this recognition of 'for someone' or 'self-awareness' amount to?

Here's how Uriah Kriegel (2009, 1) analyzes conscious experiences –

> "When I have a conscious experience of the blue sky, there is something it is like for me to have the experience. In particular, there is a bluish way it is like for me to have it. This 'bluish way it is like for me' constitutes the phenomenal character of my experience. […] The bluish way it is like for me has two distinguishable components: (i) the bluish component and (ii) the for-me component."

According to this approach, then, the 'for-me-ness' or 'subjective character' of a conscious experience is conceived of as a further component or aspect of experience. This is consistent with Dan Zahavi's phenomenological idea of the self as an experiential dimension. He suggests replacing the traditional "subject of experience" with the phrase "subjectivity of experience". On Zahavi's view the self is present to itself precisely and only when worldly engaged. According to Zahavi this is captured by the phrase "subjectivity of experience" whereas the phrase

"subject of experience" might suggest that "the self is something that exists apart from, or above, the experience and, for that reason, is something that might be encountered in separation from the experience or even something the experience may occasionally lack" (Zahavi, 2008, 126).

The limitation of this approach is that it leaves out any metaphysical discussion of a conscious subject or self. As Martine Nida-Rümelin points out "… 'for-me-ness' is conceived as being on the 'same metaphysical level', so to speak, as the phenomenal character of experience. However, what 'for-me-ness' refers to is the metaphysical structure of something being phenomenally given *to someone*" (Nida-Rümelin, 2015, 13).

Nida-Rümelin calls this primary metaphysical structure 'basic intentionality', which expresses the idea that "the metaphysical structure of experiences" is such that "any experience requires *a subject to whom something is phenomenally given"* (Nida-Rümelin, 2014, 2). This is not a fact about the world that we might discover: rather it is a fact about "the world as I found it" prior to any particular experience. And in having any experience we are necessarily aware of that structure. It is something we know by being. As Nida-Rümelin puts it –

> "We are aware of that structure by being the subject involved in the experience, by being the one to whom something is given. Furthermore, by being aware of that structure, by being aware, in every experience, of its basic intentionality, we are aware of ourselves as the one to whom something is phenomenally given" (Nida–Rümelin, 2014, 8).

On this view then, an awareness of basic intentionality, as it is present in every experience, does not require a concept of oneself or a general concept of what it is to be a subject of experience. Rather, it is there before we start to reflect upon what is given in experience or upon the question about who (or what kind of individual) it is to whom something is given. As Nida-Rümelin argues – "That kind of awareness is there, not on the basis of experiencing, but in the act of experiencing itself " (Nida-Rümelin, 2014, 9). We know it by being the experiencing person.

8.1.3 BEING ME

Being an experiencing person involves knowing ourselves as a unique individual. As Thomas Nagel (1986) describes it, we each have an ineliminable sense of 'being me' that cannot be captured in any objective account of the world. He argues that a centerless account of the world may contain an account of a person called Thomas Nagel including all his features and characteristics, but it would not include the fact that he is that person – "how can a particular person be me?"

> "Given a complete description of the world from no particular point of view, including all the people in it, one of whom is Thomas Nagel, it seems on the one hand that something has been left out, something absolutely essential remains to be specified, namely which of them I am. But on the other hand there seems no room in the centerless world for such a further fact: the world as it is from no point of view seems complete in a way that excludes such additions; it is just the world, and everything true of TN is already in it" (Nagel, 1986, 54-55).

Objective accounts of the world seem to be complete; but from the way we experience the world there seems to be something essential that is left out.

Further, Nagel argues that from our sense of 'being me' another question arises –

"how can I be *merely* a particular person? … how can I be anything as specific as a particular person in the world at all – any person?" (Nagel, 1986, 55).

He argues that this second question arises from something about the idea of 'I". "It can seem that as far as what I really am is concerned, any relation I may have to TN or any other objectively specified person must be accidental and arbitrary. …that 'I am TN' insofar as it is true, is not an identity but a subject-predicate proposition" (Nagel. 1986, 55).

Thus Nagel argues that "'being me' … seems undeniably to be a further truth, in addition to the most detailed description of TN's history, experience, and characteristics" (Nagel, 1986, 56). And that this "… provides a clear

example of the ineliminability of the indexicals from a complete conception of the world, and it also reveals something about each of us" (Nagel, 1986, 57). A complete conception of the world must thus include the fact of 'being me'. As Wolfgang Fasching (2009) puts it – "That I am me (the person that I am) is a "further truth" over and above *all* objective truths (be they physical or mental)". What it reveals about each of us is what Mait Edey (2009) terms "the subjective insight" – realizing for yourself that you are conscious. Edey argues that we have a primary realization of ourselves as conscious (aware) subjects before any identification of ourselves with any object in the world – "for me to realize that I am does not in itself tell me what I am, beyond awareness. For me to wonder who or what I am implies that I have already realized that I am. I may not be sure who or what I am, but I do already know, and know for certain, that I am."

What it reveals about each of us, then, is the primary fact about our selves as being an independent 'me' – a unique individual subject of experiences, a source of action, and the center of an individual perspective on the world. We know this 'me' by being it: I know myself by being me.

8.2 OBJECTIVITY, EXISTENTIAL PRIMACY AND ONTOLOGY

8.2.1 OBJECTIVITY

Although our primary sense of ourselves is as a unique individual, Nagel argues that we also have another aspect or capacity - the capacity to transcend our particular point of view and to conceive of the world as a whole. Nagel makes clear however that this aspect or capacity is, in some sense secondary to our primary sense of 'being me'. He describes it as "only part of the point of view of an ordinary person, and its objectivity is developed to different degrees in different persons and at different stages of life and civilization" (Nagel, 1986, 63).

He describes how this capacity develops in a process of 'imagination' that involves 'stepping back' in stages -

- The basic step involves "conceiving the world as a place that includes the person I am within it, as just another of its contents – conceiving myself from the outside, in other words."
- next – "conceiving from outside all the points of view and experiences of that person and others of his species,"
-and then – "considering the world as a place in which these phenomena are produced by interaction between these beings and other things" (Nagel, 1986, 63).

Nagel argues that this imaginative capacity for 'stepping back' is the beginning of science. And he stresses that this process, and therefore science, is an achievement of the individual 'me' - "…it is I who have done this stepping back, not only from an individual viewpoint but from a specific type of viewpoint" (Nagel, 1986, 63).

Therefore, as I argued in Chapter 5 science should be understood as an abstraction from our lived reality. As Evan Thompson puts it – "direct experience is primary and science secondary" (Thompson, 2015, 100). According to Thompson the scientific method gives us no direct or independent access to consciousness. As David Chalmers (1998, 220) puts it –"we have no consciousness meter". This is because consciousness isn't the kind of property that can be measured directly. Any meter records a publicly observable property such as temperature and assigns some magnitude to it. However, consciousness isn't publicly observable or quantifiable; rather it's subjective and qualitative. As Thompson explains it –

> " Whenever we construct a meter to measure something, we must abstract away from any associated subjective and qualitative aspects, as we do when we abstract away from felt warmth in order to measure temperature. It simply makes no sense to suppose we could use a meter to measure what we're required to abstract away from it in order to make any meter function" (Thompson, 2015, 97).

Of course scientists can, and do, measure what they believe to be publicly observable signs or indicators of the presence of consciousness. So, for

example, the Bispectral Index is an EEG (electroencephalogram) measure used during anesthesia to monitor a patient's level of consciousness and thus the depth of anesthesia. But there is no scientific way to bypass the signs and indicators and go straight to consciousness itself. This is what Ned Block (2008) calls the "methodological puzzle" in the science of consciousness –

> "how can we find out whether there can be conscious experience without the cognitive accessibility required for reporting conscious experience, since any evidence would have to derive from reports that themselves derive from that cognitive access?"

Block's puzzle points to the fact that science has no access to experience that's independent of consciousness; the scientific method always necessarily involves using and relying on consciousness itself in activities such as perceptual observation, asking questions, formulating hypotheses, analyzing data and communicating results.

Evan Thompson therefore concludes that we are not in a position to get a complete world view, just from science –

> "…there's no way to stand outside consciousness and look at it, in order to see how it fits with the rest of reality. Science always moves within the field of what consciousness reveals; it can enlarge this field and open up new vistas, but it can never get beyond the horizon set by consciousness. In this way, direct experience is primary and science secondary" (Thompson, 2015, 100).

This is in line with my argument in Chapter 5 that objective science is an important, but not our only route towards the ultimate nature of reality; we should also make use of our knowing of selves as conscious subjects.

8.2.2 EXISTENTIAL PRIMACY

Because science gives us no direct or independent access to consciousness, Thompson argues that consciousness has an epistemological or methodological primacy in any scientific investigation (Thompson, 2015, 100).

More fundamentally though, consciousness is existentially primary: it is something we live not something we have. As I put it in Chapter 1, consciousness is a primary fact about the way we find ourselves living; 'the world as I found it' involves 'I' finding – becoming conscious of – a world. And, Nagel's ineliminable sense of 'being me', is of 'me' as a conscious subject. Thompson writes – "We don't have consciousness in any sense that would allow us to lose it without ceasing to be what we are. Consciousness is our way of being, and it cannot be objectified, that is, treated as just another kind of object out there in the world, because it is that by which any object shows up for us at all" (Thompson, 2015, 100).

The observation that consciousness is existentially primary, therefore, is not a metaphysical theory about consciousness, such as idealism, dualism or panpsychism, which are about the ontological status of consciousness as contrasted with the ontological status of matter. Rather it is about paying attention to our own lived experience, which is the precondition of developing any theory whatsoever.

Michel Bitbol (2008) explains the peculiar all-pervasive primacy of consciousness by paraphrasing Wittgenstein's remark on sensations: "[Consciousness] is not a *something*, but not a *nothing* either!" (Wittgenstein, 1983, 304). Bitbol argues that consciousness as experience *is not something,* in that it is neither an object, a property, a phenomenon nor even an illusion. But it is not *nothing,* because although consciousness "is not something that we *have,* [rather] it identifies with what we *are* in the first place. It is not something that can be known or described by us in the third person as if we were separated from it; but it is what we dwell in and what we live through *in the first person.*" In this sense then Bitbol argues that consciousness is *existentially* primary.

In everyday life, then, we don't have to proceed like scientists who have to infer the inner presence of consciousness on the basis of outer signs, in a process of inference to the best explanation. In our own case we have direct knowing of our own consciousness, and before any reflection (including that of science), we already implicitly recognize each other as conscious on the basis of empathy. The phenomenological understanding of empathy is as the direct perception of another being's actions and gestures as expressive embodiments of consciousness. As Thompson explains it –

> "We don't see facial expressions, for example, as outer signs of an inner consciousness, as we might see an EEG pattern; we see joy directly in the smiling face or sadness in the tearful eyes. Moreover, even in difficult or problematic cases where we're forced to consider outer criteria, their meaningfulness as indicators of consciousness ultimately depends on and presupposes our prior empathic grasp of consciousness" (Thompson, 2015, 99).

The existential priority of consciousness then is the all-pervasive feature of the world as we find it before any reflection whatsoever.

8.2.3 THE ONTOLOGICAL STATUS OF CONSCIOUSNESS

But Evan Thompson (2015, 102) argues that we can't infer from the existential or epistemological primacy of consciousness that consciousness has *ontological* primacy in the sense of being the primary reality out of which everything is composed or the ground from which everything is generated. This doesn't follow because –

(a) That the world as we know it is always a world for consciousness, doesn't logically entail that the world is made out of consciousness.

(b) The testimony of our direct experience is of the contingency of our consciousness on the world, specifically on our living body and environment.

As Thompson sees it there are two points of view that we need to hold together without privileging one over the other –

- (1) we can never step outside consciousness to see how it measures up to something else
- (2) consciousness never appears or shows up apart from some context of embodiment.

Thompson borrows a simile from the Buddha to express our unresolved experience of consciousness and embodiment as being like "two sheaves of reeds [that] stand leaning against each other" (Bodhi, 2002, 608-609). Thompson writes –

> "We can never step outside consciousness, but consciousness always shows up contingent on our embodiment, and we can never step outside our embodiment, but our embodiment always shows up contingent on consciousness" (Thompson, 2015, 102).

Similarly, Martine Nida-Rümelin argues that the ontological issue about the nature of conscious beings, and the phenomenological issue about the way in which we are aware of ourselves in experience, turn out to be intimately related in a way that excludes their successful independent analysis –

> " The phenomenological thesis that we are in everyday experience pre-reflectively and pre-conceptually aware of ourselves *as experiencing subjects* can only be explained in a clear manner by an account of what experiencing subjects are. On the other hand, ontological claims about the nature of experiencing subjects, their capacity to be in the subject position of an event exhibiting basic intentionality and their capacity to ground synchronic and diachronic unity, can only be understood and can only be convincingly motivated on the basis of phenomenological insights concerning the way in which we are in every moment of our conscious life pre-reflectively aware of ourselves as experiencing beings" (Nida-Rümelin, 2017, 30).

In trying to understand our world, then, our view is inevitably limited to some extent by our situation; we cannot get outside the dualities of consciousness and embodiment or phenomenology and ontology.

And as Nagel warns, objectivity can be both underrated and overrated, sometimes by the same persons –

"It is underrated by those who don't regard it as a method of understanding the world as it is in itself. It is overrated by those who believe it can provide a complete view of the world on its own, replacing the subjective views from which it has developed" (Nagel, 1986, 5).

As Nagel sees it "Instead of a unified world view, we get the interplay of these two uneasily related types of conception [internal and external or subjective and objective], and the essentially incompatible effort to reconcile them. The transcendent impulse is both a creative and a destructive force" (Nagel, 1986, 4).

The limit encountered by the pursuit of objectivity appears precisely when it turns back on the self and tries to encompass subjectivity in its conception of the real. Thus Nagel argues that -

> "The fundamental idea behind both the validity and the limits of objectivity is that we are small creatures in a big world of which we have only very partial understanding, and that how things seem to us depends both on the world and on us" (Nagel, 1986, 5).

Our limited ontological knowledge leaves us in a situation of subject and object duality but this is not a matter of dividing the world; rather it is a matter of recognizing our own situation and the view we might have from this location.

8.3 THE VIEW FROM HERE

Being me is the locus of all experience. A first-personal bodily experience that we are is the primary feature of the way we find ourselves living and the source of all knowledge and understanding. Any possible experiences are achievements of an individual me, and any universal claim about reality must include the subject that produces it.

Being me is therefore the central feature of the world as we find it, and the only locus from which we can construct our view of reality. As Frank, Gleiser and Thompson put it –

> "The best we can aim for is to construct a scientific cosmological account that is consistent with what we can measure and know *from inside.* The account can't ever be a final or complete description of cosmic history. Rather, it must be an on-going, self-correcting narrative."

On this view, what we should aim for is 'a science nourished by [a] sensibility for humanity'.

Although this may sound deflationary, it can also be liberating. Although 'the world as I found it' is always already from an individual unique perspective, it is the perspective of a 'me' that is inextricably part of a social and wider world, that we can seek to understand. Being me therefore involves an understanding of ourselves as 'both as an expression of nature and as a source of nature's self-understanding" (Frank, Gleiser and Thompson, 2019). Knowing ourselves gives us knowledge about the world: and our capacity for explanatory judgment instantiates nature's understanding of itself.

CONCLUSION

The central claim of this dissertation is that our knowing by being of ourselves is as individual conscious selves that are bodily subjects with a first-person perspective. Although we find ourselves with others in a wider world, 'being me' is the existentially primary fact that we know by being. This knowledge is inconsistent with any exclusively third-person theory that claims to have universal scope.

I take the existence of unique individual selves, with others in a wider world to be the primary facts about the world as we find it. This claim is not held with any certainty that it is the ultimate theory of everything. We cannot assume that we can ever know this. It is however based on the evidence of our own experience, which is what we know with the greatest certainty. In the absence of any reason to think otherwise, it is reasonable to require that any account of the way things are must do justice to the way things seem to us.

I reject physicalism as an adequate universal ontological theory because it fails to explain the nature of physical matter, which it claims is fundamental, and fails to account for consciousness, which is necessary to assert any theory whatsoever. Joel Walmsley (2012, 171) describes physicalism as "the thesis that we are (just) matter in motion", and argues that to deny physicalism is "to say that we are something *more*" – a claim that he describes as "simply untenable". A conclusion of this dissertation is that physicalism as a universal metaphysical theory is untenable because it gets things backwards: it attempts to explain what is directly given to us (conscious experience) from what is given indirectly (whatever physics decides is fundamental). The claim is not that we are something 'more', but that human experience is prior to any theory whatsoever.

Ladyman and Ross's (2007, 220) rainforest realism is an improvement on traditional physicalism in that it allows for a non-reductive ontology that asserts the reality of patterns at many different scales. However it too is a wholly third person account that has no adequate place for the observer of patterns.

The popularity of these universal aspiring metaphysical theories is based on a misplaced belief that science can give us a complete, objective description of reality – a God's eye view of nature - distinct from us, and our perception of it. Science is claimed to be our only method of discovering the objective character of the world.

However, I claim that the primary fact of 'being me' is the locus of, and frame for, all our knowledge, including science and any metaphysical theory about what we are. All our knowledge is accessed from a first-person perspective and what we call 'reality' is inextricable from our experience.

This has meta-scientific implications about how we understand what science is and what it can tell us. Science should be seen as an on-going self-correcting exploration of our world from within, rather than as direct access to a mind independent reality that will culminate in a theory of everything. This approach also has intra-scientific implications, in that scientific models are always bound to our observations and actions, and circumscribed in their application.

This reframing of science does not imply that particular sciences or theories do not access aspects of the world. Rather, my claim is that the overall project of science is properly understood as a human project, with a certain methodology, and there is always a close connection between our experience and the world. As Merleau-Ponty (1964, 27) put it –

> "… a physics that has learned to situate the physicist physically, a psychology that has learned to situate the psychologist in the socio-historical world, have lost the illusion of the absolute view from above: they do not only tolerate, they enjoin a radical examination of our belongingness to the world before all science."

Although in this dissertation I have focused on scientistic physicalism, the rejection of universal theories that neglect their first personal starting point could be applied also to other areas such as religious doctrines that claim to have answers to universal questions, and meditative practices that claim we can and should dissolve our individuality into a universal oneness. Universal religious doctrines are analogous to scientistic theories of everything that

claim a degree of knowledge that is not available to us. Universal oneness is neither possible nor desirable for the unique kind of self that we are.

My stress on the significance of individuals at the ontological level is not related in any way to, and should not be used as an argument for, individualism at the political or economic level. On the contrary, I argue that we are individuals always interdependent with others and the world. This interdependence should be central to how society and the economy are organized. On the level of social ontology however, groups and institutions are ultimately to be understood in terms of, and are the responsibility of, the individuals that constitute them.

The kind of self that we know ourself to be is best understood as "an expression of nature and as a source of nature's self-understanding" (Frank, Gleiser and Thompson, 2019). This exemplifies what Merleau-Ponty termed "our belongingness to the world", and that "our body is our living bond with nature" (Merleau-Ponty, 1964, 27). Our direct knowing by being of ourselves as experiencing bodily subjects expresses important features of nature that we can come to understand from the way we find ourselves living. Our experience of ourselves as agents and knowers reveals our cosmic significance and our responsibility as a source of nature's self-understanding.

REFERENCES

Albahari, M. 2006. *Analytical Buddhism: The Two-Tiered Illusion of Self.* New York: Palgrave Macmillan.

Alexander, Samuel. 1920. *Space, Time, and Deity*, vol. II, Peter Smith, 1979.
---. 1924. 'Preface to New Impression', *Space, Time and Deity* vol. 1. London: MacMillan.

Ames, Roger T. 2011. *Confucian Role Ethics: A Vocabulary.* Hong Kong: The Chinese University Press.

Ames, Roger T. and Rosemont, Henry. 1998. *The Analects of Confucius. A Philosophical Translation*. New York: Ballantine Books.

Arendt, Hannah. 1958. *The Human Condition.* University of Chicago Press.

Baker, Lynne Rudder. 1987. Saving Belief: A Critique of Physicalism. Princetown: Princetown University Press.
---. 2007. "Our Place in Nature: Material Persons and Theism". PhilArchive copy v1: https://philarchive.org/archive/BAKOPIv1.
---. 2013. *Naturalism and the First-Person Perspective.* Oxford: Oxford University Press.

Banegas, Diane et al. 2007. *At Discovery's Horizon: Report of the Task Force on US LHC Communication.* Tech. Rep., Natioal Science Foundation.

Barash, David P. 2018. "Anthropic Arrogance". *Aeon* 18-9-2018.

Bargh, John. 2017. *Before you know it: The unconscious reasons we do what we do.* William Heinemann: London.

Barrow, John D. 2007. *New Theories of Everything.* Oxford: Oxford University Press.

Barrow, John D. and Tipler, Frank J. 1986. *The Anthropic Cosmological Principle.* Oxford: Oxford University Press.

Berlin, Isaiah. 1979. *Against the Current: essays in the History of Ideas*. London: Pimlico.

Block, N. 2007. *Consciousness, Function, and Representation.* Cambridge, MA: MIT Press.

Block, Ned, and Cynthia MacDonald. 2008. "Consciousness and Cognitive Access". *Proceedings of the Aristotelian Society* CVIII, Part 3 (2008): 289-316.

Bitbol, Michel. 2002. "Science as if situation mattered" published in *Phenomenology and Cognitive Science,* 1, 181-224, 2002.
---. 2008. "Is Consciousness primary?" *NeuroQuantology,* vol. 6, no 1, 53-72, 2008.

Bitbol, Michel and Luisi, Pier-Luigi. 2011. "Science and the Self-Referentiality of Consciousness", *Journal of Cosmology,* 2011, Vol.14, 207-223.

Bitbol, Michel & Petitmengin, Claire. 2013. "On the possibility and reality of introspection", *Kairos. Journal of Philosophy and Science* 6: 2013.

Bohm, David. 1980. *Wholeness and the Implicate Order.* Ark Paperbacks.

Bohr, N. 1934. *Atomic Theory and the Description of Nature* 18, Cambridge University Press.Bolander, Thomas, "Self-Reference", *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2015/entries/self-reference/>.

Bratman, M. 1993. "Shared Intention", *Ethics,* 104: 07-113.
---. 2014. *Shared Agency. A Planning Theory of Acting Together.* Oxford: Oxford University Press.

Brentano, Franz. 1874. Psychology from an Empirical Standpoint, transl. by A.C. Rancurello, D.B. Terrell, and L. McAlister, London: Routledge, 1973. (2nd ed., intr. by Peter Simons, 1995).

Brewer, M.B. 1991. "The Social Self: On being the same and different at the same time" *Personality and Social Psychology Bulletin* 17, 475-482.

Broad, C.D. 1925. *The Mind and Its Place In Nature.* London: Routledge and Kegan Paul.

Bruge, T. 1997. 'Two Kinds of Consciousness' In N. Block, O. Flanagan, and G. Guzeldere (eds.), The Nature of Consciousness: Philosophical Debates, pp. 427-433. Cambridge, MA: MIT Press.

Bufacchi, Vittorio and Ouyang, Xiao. 2017. "Hens, Ducks, & Human Rights in China". *Philosophy Now.* Issue 118, February – March 2017.

Burnet, F. M. 1961. Immunological recognition of self. (Nobel Lecture) *Science* 133, 307-311.

Burnet, F.M. & Fenner, F. 1949. *The Production of Antibodies* (2nd. ed.), Melbourne, Macmillan.

Canguilhem G (1994a) [1963] La constitution de la physiologie comme science. In: *Etudes d'histoire et de philosophies des sciences concernant les vivants et la vie*. Vrin, Paris, pp. 226–273.

Carroll S. 2010. “The laws underlying the physics of everyday life are completely understood”. *Cosmic Variance*, September 23.

Carter, Brandon. 1974. “Large Number Coincidences and the Anthropic Principle in Cosmology” in M.S. Longair (eds.) *Confrontation of Cosmological Theories with Observational Data* 291-298.

Cartwright, Nancy. 1983. *How the Laws of Physics Lie.* Oxford: Oxford University Press.
---. 1999. *The Dappled World: A Study of the Boundaries of Science.* Cambridge: Cambridge University Press.

Chalmers, David J. 1996. *The Conscious Mind: In Search of a Fundamental Theory.* Oxford: Oxford University Press.
---. 1998. “On the Search for a *Neural* Correlate of Consciousness”. in Stuart R. Hameroff et al., eds., *Towards a Science of Consciousness II,* Cambridge MA: MIT Press.
---. 2016. “The Combination Problem for Panpsychism”. In T. Alter and Y. Nagasawa (eds.), *Panpsychism.* Oxford: Oxford University Press.

Chaitin, Gregory. 1975. “Randomness and Mathematical Proof,” *Scientific American.* CCXXXII (1975): 47-52.

Church, A. 1936. *Mathematical Logic.* Princetown University.

Clark, A. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36, 181–204.
---. 2016. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press, NY.
---. 2017. “A nice surprise? Predictive processing and the active pursuit of novelty”. *Phenomenology and the Cognitive Sciences.* 17, 521-534.

Ciaunica, Anna & Charlton, Jane. 2018. “When the self slips”. *Aeon*. 21 June 2018.

Davies, Paul. 2006. *The Goldilocks Enigma*. Penguin Books.

Dennett, Daniel. 1991. *Consciousness Explained*. Boston, MA: Little, Brown and Co.
---. 1991a. “Real Patterns”. *The Journal of Philosophy,* Vol.88. 1. 9, Jan.1991, pp. 27-51.
---. 1996. *Darwin’s Dangerous Idea: Evolution and the Meaning of Life.* New York: Simon & Schuster.
---. 2016. “Illusionism as the Obvious Default Theory of Consciousness”. *Journal of Consciousness Studies*, 23, No. 11-12, 2016, pp.65-72.
---. 2017. *From Bacteria to Bach and Back: The Evolution of Minds.* Allen Lane.

Deikman, Arthur J. 1996. “I = Awareness” *Journal of Consciousness Studies,* 3 (4), 350-356.

Descartes, R. 1986. *Meditations on First Philosophy.* Cambridge: Cambridge University Press.

Deutsch, David. 2011. *The Beginning of Infinity: Explanations that Transform the World.* London: Penguin Books.

Dicke, R. H. 1957. “Principle of Equivalence and Weak Interactions,” *Rev. Mod. Phys.*, **29**, 355 (1957).

Di Paolo, Ezequiel, and Thompson, Evan. 2014. “The Enactive Approach”. in Lawrence Shapiro, ed., *The Routledge Handbook of Embodied Cognition* (Routledge Press).

Dupré, John. 1993. *The Disorder of Things.* Harvard University Press.

Dyson, Freeman. 1953. "Field Theory". *Scientific American.* April 1953.

Edey, Mait. 2009. "Subject and Object". Chapter 8 of *Nature and Self: Reframing the Human Predicament* (A work in progress) Copyright © 2009 by Mait Edey. *ANTIMATTERS* 3 (1) 2009.

Ellemers, Naomi. 2012. "The Group Self", *Science,* Vol. 336, Issue 6083, 848-852.

Ellis, George F. R. 2005. "Physics and the Real World". *Physics Today.* 58,7,49.

---. 2008. "Issues in the Philosophy of Cosmology". *Studies in History and Philosophy of Science – Part B - Studies in History and Philosophy of Modern Physics.* 46(1) March 2006.

Epicurus. 1964. *Letters, Principle Doctrines and Vatican Sayings.* Tr. R. M. Geer, Indianapolis and New York: Bobbs-Merrill.

Everett, Hugh. 1955. *The Theory of the Universal Wavefunction*, Thesis, Princeton University (1956, 1973) pp. 1–140.

Evans, Jonathan St. B. T. 2003. "In Two Minds: Dual-Process Accounts of Reasons" *Trends in Cognitive Science* 7 (2003): 454-459.

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

---. 2016. "The Nonplurality of the I. On the Question of the Ultimate Subject of Experience" *Journal of Consciousness Studies* 23/1-2, 2016.

Fingarette, Herbert. 1983. "The Music of Humanity in the Conversations of Confucius," in *The Journal of Chinese Philosophy*, vol. 10, 1983.

Flanagan, Owen. 1992. *Consciousness Reconsidered.* Cambridge, Mass.: MIT Press.
---. 2011. *The Bodhisattva's Brain. Buddhism Naturalized.* Cambridge, Mass.: MIT Press.

Frank, Adam. 2017. "Minding Matter", *Aeon,* 13-3-2017.

Frank, A., Gleiser, M. & Thompson, E. 2019. "The Blind Spot". *Aeon,* 8-1-2019.

Frankish, Keith. 2016. "Illusionism as a Theory of Consciousness", *Journal of Consciousness Studies,* 23, No. 11-12, 2016, pp. 11-39.
---. 2018. "Why panpsychism fails to solve the problem of consciousness", *Aeon* 27th October 2018.

Friston, K. 2010. The free-energy principle: A unified brain theory? Nature Reviews. *Neuroscience*, 11(2), 127–138.

Fuchs, C.A. & Schack, R. 2013. "Quantum – Bayesian coherence". *Rev. Mod. Phy.* 85, 1683-1715 (2013)

Fuchs, C.A. 2016. "On Participatory Realism" in *Information and Interaction: Eddington, Wheeler, and the Limits of Knowledge.* Ian T. Durham and Dean Riches (Eds.). Springer.

Fuchs, Thomas. 2017. "Self across time: the diachromic unity of bodily existence". *Phenom Cogn Sci* (2017) 16: 291-315.

Galileo Galilei. 1623. *Il Saggiatore* (in Italian) (Rome, 1623); *The Assayer*, English trans. Stillman Drake and C. D. O'Malley, in The Controversy on the Comets of 1618 (University of Pennsylvania Press, 1960).

Gallagher. Shaun. 1986. "Body image and body schema: a conceptual clarification. *The Journal of Mind and Behavior* 7: 541-554.
---. 1995. "Body schema and intentionality". In J.L. Bermudez, A. Marcel, and N. Eilan, eds., *The Body and the Self.* Pp. 225-244. Cambridge. MA: MIT Press.
---. 2012. *Phenomenology,* Palgrave Philosophy Today.

Gallagher, Shaun and Zahavi, Dan. 2008. *The Phenomenological Mind,* London and New York: Routledge.
---. 2010. "Phenomenological Approaches to Self-Consciousness", *The Stanford Encyclopedia of Philosophy* (Winter 2010 Edition), Edward N. Zalta (ed.)

Gardner, Martin. 1986. "WAP, SAP, PAP, and FAP". *The New York Review of Books.*23, No. 8 (May 8, 1986): 22-25.

Garfield, J. 2015. *Engaging Buddhism: Why it matters to philosophy.* Oxford: Oxford University Press.
---. 2016. "Illusionism and Givenness", *Journal of Consciousness Studies*, Volume 23, Numbers 11-12, 2016, pp. 73-82(10).

Gilbert, Margret. 1987. "Modelling Collective Belief", *Synthese,* 73, 185-204.

Gödel, Kurt. 1934. *On Undecidable Propositions of Formal Mathematical Systems.* Princeton University.
---. 1962. *On Formally Undecidable Propositions,* New York: Basic Books.

Godfrey-Smith, Peter. 2016. *Other Minds.* London: William Collins Books.

Goff, Philip. 2017. *Consciousness and Fundamental Reality.* Oxford: Oxford University Press.

Gould S.J., Lloyd E.A. 1999. Individuality and adaptation across levels of selection: How shall we name and generalize the unit of Darwinism? *Proc Natl Acad Sci* 96:11904–11909.

Hall, David and Ames, Roger. 1987. *Thinking Through Confucius.* Amherst, NY: SUNY Press.

Harris, Rendel. 1906. "The Cretans always liars". *The Expositor Seventh Series* 2:305-17.

Hasker, William. 1999. *The Emergent Self.* New York: Cornell University Press.

Hawking, Stephen. 1995. From an interview with Ken Campbell on the 1995 show *Reality on the Rocks: Beyond Our Ken.*

Hempel, Carl Gustav. 1969. "Reduction: Ontological and linguistic facets", in S. Morgenbesser, P. Suppes, and M. White (eds.), *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*, New York: St. Martins Press, pp. 179–199.

Henry, Aaron, & Thompson, Evan. 2011. "Witnessing From Here: Self-awareness from a bodily versus embodied perspective." In *The Oxford Handbook of the Self.* Edited by Shaun Gallagher. Oxford: Oxford University Press.

Heidegger, Martin. 1973. *Being and Time.* Oxford: Basel Blackwell.

---. 1979. *Prolegomena zur Geschichte des Zeitbegriffs*. Gesamtaugabe Band 20. Frankfurt am Main: Vittorio Klostermann.
---. 2001. Gesamtausgabe Band 27: *Einleitung in die Philosophie.* Frankfurt am Main. Vittorio Klostermann, 2001.

Hellman, G. P. and Thompson, F. W. 1975. "Physicalism: Ontology, determination and reduction. *Journal of Philosophy* 72 (October) 551-564.

Hobbes, T., 1651, *Leviathan*, in E. Curley (ed.), *Leviathan, with selected variants from the Latin edition of 1668*, Indianapolis: Hackett, 1994.

Hofstadter, Douglas R. 1979. *Gödel, Escher, Bach: An Eternal Golden Braid,* Penguin Books.

Horst, Steven. 2014. "Beyond Reduction: From Naturalism to Cognitive Pluralism" *Mind and Matter.* Vol. 12, Issue 2, 2014.

Howell, Robert J. 2010. "Subjectivity and the Elusiveness of the Self" *Canadian Journal of Philosophy*. Vol. 40. No. 3, September 2010.

Hohwy, Jakob. 2013. *The predictive mind*. Oxford: Oxford University Press.

Hohwy, Jakob, and John Michael. 2017. "Why Should Any Body Have a Self?" *PsyArXiv*. April 20. doi:10.31234/osf.io/fm4cr.

Huebner, Bryce. 2013. *Macrocognition: A Theory of Distributed Minds and Collective Intentionality.* Oxford: Oxford University Press.

Hume, David. 1739/1975. *A Treatise of Human Nature*. Oxford: Oxford University Press.

Husserl, Edmund. 1913/1931. *Ideas: General Introduction to Pure Phenomenology*, London: George Allen & Unwin.

---. 1931/1960. *Cartesian Meditations: An Introduction to Phenomenology.* Trans. Q.Cairns, Den Haag; Martinus Nijhoff.
---. 1956. *Erste Philosophie (1923/24). Erster Teil: Kritische Ideengeschichte, ed. Rudolf Boehm. Den Haag: Nijhoff (Husserliana VII).*
---. 1959. *First philosophy* (1923/24). Part two: theory of phenomenological reduction. Edited by Rudolf Boehm. 1959. xlii + 592 pp.
---. 1962. *Phänomenologische Psychologie.* Husserliana IX. Ed. By W.Biemel, Den Haag: Martinus Nijhoff.
---. 1936/1970. *The Crisis of European Sciences and Transcendental Phenomenology*. Northwestern University Press. Evanston.
---. 1969. *On the phenomenology of inner time consciousness (1893-1917).* Edited by Rudolf Boehm. Reprint of the 2nd verb. Edition. 1969.
---. 1973(a). On the phenomenology of intersubjectivity. Texts from the estate.
Second part: 1921-1928. Edited by Iso Kern. 1973. xxxvi + 624 pp.
---. 1973(b). On the phenomenology of intersubjectivity. Texts from the estate.
Third part: 1929-1935. Edited by Iso Kern. 1973. lxx + 742 pp.
---. 1984. Logical investigations . Volume two: Studies on phenomenology and theory of knowledge. Edited by Ursula Panzer. 1984. 1024 pp. in two volumes.
---. 1989. *Ideas pertaining to a pure phenomenology and to a phenomenological philosophy.* Kluwer Academic Publishers.

Hume, David. 1739/1975. *A Treatise of Human Nature*. Oxford: Oxford University Press.

Hutto, D. and Myin, E. 2017. *Evolving Enactivism*. Cambridge, MA: MIT Press.

Ismael, J. 2006. "Saving the Baby: Dennett on Autobiography, Agency, and the Self. *Philosophical Psychology.* Vol. 19, No. 3, June 2006, pp.345-360.

Jackson, Frank. 1982. "Epiphenomenal Qualia" *The Philosophical Quarterly,* Vol. 32, No 127 (Apr. 1982), pp.127-136.

James, William. 1890. *Principles of Psychology,* New York: Dover, 1950.

Jaworski, William. 2016. *Structure and the Metaphysics of Mind.* Oxford: Oxford University Press.

Jonas, Hans. 1966. *The Phenomenon of Life*. Northwestern University Press.
---. 1968. "Biological foundations of individuality" *International Philosophical Quarterly* 8: 231-251.
---. 1996. *Mortality and Morality: A Search for the Good after Auschwitz.* Evanston, II: Northwestern University Press.

Joseph, Rhawn. 1982. "The Neuropsychology of Development. Hemispheric Laterality, Limbic Language, the Origin of Thought", *Journal of Clinical Psychology,* 44 4-33.

Kahneman, Daniel. 2011. *Thinking Fast and Slow.* New York: Farrar, Straus and Giroux.

Kane, Robert. 2005. "Some Neglected Pathways in the Free Will Labyrinth" in *The Oxford Handbook of Free Will.* Oxford: Oxford University Press.

Kant, I. 1784/1981. "Idea of a Universal History from a Cosmopolitan Point of View," in *On History,* by Immanuel Kant, translated by Lewis White Beck. Bobbs-Merrill Library of Liberal Arts.
---. 1781/1797. *Critique of Pure Reason.* Trans, Marcus Weigelt. Penguin Books.
---. 1790/1987. *Critique of Judgment.* Trans, W.S.Pluhar. Indianapolis, IN: Hacket Publishing Company.

Kenny, Anthony. 2010. *A New History of Western Philosophy.* Oxford: Oxford University Press.

Kim, Jaegwon. 2005. *Physicalism, or Something Near Enough.* Princeton and Oxford: Princeton University Press.
---. 2010. *Essays in the Metaphysics of Mind.* Oxford: Oxford University Press.

Kitcher, P. 1981. Explanatory unification. *Philosophy of Science* 48: 507-31.

Klausen, Søren. 2019. "The Essential Tension in Phenomenal Consciousness". *Philosophical Papers.*

Knausgård, Karl Ove. 2018. *The End. My Struggle: Book 6.* Translated from Norwegian by Marten Aithen and Don Bartlett. London: Harvill Secker.

Kriegal, U. 2004. "Consciousness and Self-Consciousness." *Monist* 87/2, 185-209.
---. 2009. *Subjective Consciousness: A self-Representational Theory.* Oxford: Oxford University Press.

Ladyman, J. and Ross, D. 2007. *Every Thing Must Go – Metaphysics Naturalized.* Oxford: Oxford University Press.
---. 2013. "The World in the Data" in *Scientific Metaphysics.* Edited by Don Ross, James Ladyman & Harold Kincaid. Oxford: Oxford University Press.

Lakoff, George and Johnson, Mark. 1999. *Philosophy in the Flesh*. New York: Basic Books.

Laudan, Larry. 1981. "A Confutation of Convergent Realism", *Philosophy of Science.* 48: 19-48.

Laughlin, Charles D. & Throop, C. Jason. 2009. "Husserlian Meditations and Anthropological Reflections: Towards a Cultural Neurophenomenology of

Experience and Reality." *Anthropology of Consciousness,* Vol. 20, Issue 2, pp. 130-170. (2009).

Leibniz, Wilheim Gottfried, 1686. *Discourse on Metaphysics*, Section 9 (Loemker 1969: 308).

Legrand, D. 2006. "The bodily self: the sensorimotor roots of pre-reflective self-consciousness". *Phenomenology and the Cognitive Sciences* 5: 89-118

Levine.J. 2001. *Purple Haze: The Puzzle of Consciousness.* Oxford: Oxford University Press.

Lewontin R.C. 1970. "The Units of Selection". *Annu Rev Ecol Syst* 1:1–18. doi: 10.1146/annurev.es.01.110170.000245

Lin Yutang. 2001 edn. *The Importance of Living.* Patterson Press.

Lucas, John R. 1961. "Minds, Machines and Godel". *Philosophy* 36:112-127.

Locke, John. 1689/1986. *Second Treatise on Government.* Amherst, NY: Prometheus Books.

Lockie, Robert. 2018. *Free Will and Epistemology.* New York: Bloomsbury Academic.

Lonergan, B. 1967. *Collection*, ed., F. Crowe, New York: Herder and Herder.

Lowe, E. J. 2008. *Personal Agency.* Oxford: Oxford University Press.

Luhmann, Niklas. 1986. "The Individuality of the Individual," in *Reconstruction Individualism,* edited by T.C.Heller, Moirton Sonsa, and David Wellbery. Stanford University Press, 1986.

Mackenzie, Matthew. 2011. "Enacting the Self: Buddhist and Enactivist Approaches" in *Self, No Self? Perspectives from Analytical, Phenomenological, & Indian Traditions.* Edited bu Mark Siderits, Evan Thompson, and Dan Zahavi. Oxford: Oxford University Press.

Madell, Geoffrey. 1981. *The identity of the self.* Edinburgh University Press.

Malcolm, Norman. 1968. "The Conceivability of Mechanism" *The Philosophical Review* Vol. 77, No. 1 (Jan., 1968).

Marcel, Anthony J. 2003. "The sense of agency: Awareness & ownership of action" in Johannes Roessler & Naomi Eilan (eds.) *Agency & Self-Awareness.* Oxford: Claradon Press.

Maturana, H.R. and Varela, F.J. 1973. *De máquinas y seres vivos: Una teoría de la organizacíon biologica*. Santiago, Chile: Editorial Universitaria.
---. 1980. "Autopoiesis and Cognition: The Realization of the Living"*. Boston Studies in the Philosophy of Science*, vol.42. Dordrecht: D. Reidel.
---. 1987. *The Tree of Knowledge; The Biological Roots of Human Understanding.* Boston: Shambla Press/New Science Library.

Mc Dowell, John. 2013. "The Myth of the Detached Mind" in *Mind, Reason, and Being in the World: the McDowell-Dreyfus Debate.* Ed. J.K. Schear. Routledge.

McFarland, D. 1989. *Problems of Animal Behaviour.* Harlow, Essex, UK: Longman Scientific and Technical.

McLaughlin, B. P. 1992. "The Rise and Fall of British Emergentism". In Beckermann, A., Flohr, H., and Kim, J., editors, *Emergence or Reduction: Prospects for Nonreductive Physicalism.* Berlin: De Gruyter.

Mencius. 1970. *Mencius.* Translated by D. C .Lau. Penguin Books.

---. 2009. *Mencius.* Translated by Irene Bloom. Columbia University Press.

Metzinger, T. 2003. *Being No One. The Self-Model Theory of Subjectivity.* Cambridge, MA: MIT Press.

---. 2009. *The Ego Tunnel. The Science of the Mind and the Myth of the Self.* Basic Books: New York.

Merleau-Ponty, Maurice. 1962 (1945). *Phenomenology of Perception*. London. Routledge & Kegan Paul.
---. 1963 (1942). *The Structure of Behavior.* Duquesne University Press.
---. 1964. *Le visible et l'invisible,* Paris: Gallimard.

Mermin, N. David. 2014. "Physics: QBism puts the scientist back into science". *Nature,* 421-423 (27 March 2014).

Midgley, Mary. 2014. *Are you an illusion?* Routledge.

Mill, J.S. 1843. *System of Logic.* London: Longman, Green, Reader and Dyer.

Millikan, R.G. 2004. *Varieties of Meaning: The 2002 Lean Nicod Lectures.* Cambridge, MA: MIT Press.
---. 2005. *Language: A Biological Model.* Oxford: OUP.

Moreno, Alvaro. 2016. "What makes a collection of cells an integrated individuality? *Biological Identity Conference.* Institute of Philosophy, London. 2-6-2016.

Morgan, C. Lloyd. 1923. *Emergent Evolution.* London: Williams & Norgate.

Myhill, J. 1952. "Some Philosophical Implications of Mathematical Logic". *The Review of Metaphysics* 6, 165 (1952).

Nagel, Thomas. 1974. “What Is It Like to Be a Bat?” *The Philosophical Review,* Vol.83, No. 4 (Oct., 1974), pp. 435-450.
---. 1979. “Pansychism” in *Mortal Questions* pp. 181-95. Cambridge: Cambridge University Press.
---. 1986. *The View From Nowhere.* Oxford: Oxford University Press.
---. 2012. *Mind & Cosmos.* Oxford: Oxford University Press.

Nida-Rümelin, Martine. 2012. “The Non-descriptive Individual Nature of Conscious Beings” in Georg Gasser & Mathias Stefan (eds.), *Personal Identity: complex or simple?* Cambridge University Press.
---. 2014. “Basic Intentionality, Primitive Awareness, and Awareness of Oneself” in Anne Reboul (ed.), *Mind, Values and Metaphysics. Philosophical Papers dedicated to Kevin Mulligan.* 2014.
---. 2015. “Phenomenal Consciousness and Self-Awareness”. Research Plan for a Swiss National Science Foundation Grant.
---. 2016. “The illusion of illusionism. Comment on Keith Frankish ‘Illusionism as a Theory of Consciousness’”. *The Journal of Consciousness Studies* 2016
---. 2017. “Self-Awareness”. *Review of Philosophy and Psychology* 8(1): 55-82 (2017).
---. Forthcoming. “Experiencing subjects and so-called mine-ness”. Forthcoming in M. Guillot and M. Garcia-Carpintero (eds.), *The Sense of Mineness,* Oxford University Press.

Nordhaus, Ted. 2018. “The Earth’s carrying capacity for human life is not fixed”, *Aeon* 5th July 2018.

O’Madagain, Cathal. 2012. “Group Agents: Persons, Mobs, or Zombies?” *International Journal of Philosophical Studies,* 20:2, 271-287.

Overgaard, S. and Salice, A. 2019. “Consciousness, Belief, and the Group Mind Hypothesis”. *Synthese* (1-25) 2019.

Parfit, Derek. 1987. *Reasons and Persons*. Oxford: Claradon Press.

Penrose, Roger. 1989. *The Emperor's New Mind.* Oxford: Oxford University Press.

Perry, John. 1979. "The Problem of the Essential Indexical" *Nous.* Vol. 13 No.1 (Mar. 1979) pp. 3-21.

Petitmengin, C. 2007. "Towards the Source of Thoughts". *Journal of Consciousness Studies*, 14, No. 3, 2007, pp. 54–82
---. 2017. "Enaction as a lived experience. Towards a radical neurophenomenology". *Constructivist Foundations* 12 (2): 139-147.

Pettit, P. 2010. "Groups with minds of their own". In Alvin I. Goldman & Dennis Whitcomb (eds.) *Social Epistemology: Essential Readings*. Oxford: Oxford University Press.

Plaks, Andrew H. (Translator). 2003. *Ta Hsueh and Chung Yung (The Highest Order of Cultivation and On the Practice of the Mean).* Penguin Books.

Popper, Karl. 1959. *The Logic of Scientific Discovery*. London: Hutchinson.
---. 1963. *Conjectures and Refutations*. London: Routledge and Kegan Paul.

Pradeu, Thomas. 2010. "What is an organism? An immunological answer". *History and Philosophy of the Life Sciences* 32: 247-268.
---. 2012. *The limits of the self: Immunology and biological identity.* Elizabeth Vitanza (trans.); Oxford: Oxford University Press.
---. 2013. "Immunity and the emergence of individuality" in *Groups to Individuals: Evolution and Emerging Individuality.* Edited by Frederic Bouchard and Philip Hunemam. MIT Press.

---. 2016. "Organisms or biological individuals? Combining physiological and evolutionary individuality" in *Biology & Philosophy*, November 2016, 31(6), pp. 797-817.

Pradeu, Thomas and Vivier, Eric. 2016. "The discontinuity theory of immunity" in *Science Immunology* Volume 1, Issue 1.

Proust, M. 2003. *The Way by Swann's.* translated by Lydia Davis. London: Penguin.

Quine, W.V. 1953. "On what there is" in *From a Logical Point of View*. Cambridge, Mass.: Harvard University Press. pp. 1-19 (1953)
---. 1969. *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Rees, M.J. 2003. *Our Cosmic Habitat.* Princeton and Oxford.

---. 2017. "Black holes are simpler than forests and science has its limits", *Aeon,* 1st. December 2017.

Rorty, Richard. 1979. *Philosophy and the Mirror of Nature.* Princeton University Press.

Rose, S. 1997. *Lifelines: Biology Beyond Determinism.* Oxford: Oxford University Press.

Rosemont, Henry. 2015. *Against Individualism – A Confucian Rethinking of the Foundations of Morality, Politics, Family, And Religion. .* Lexington Books.

Ross, Don. 2018. "The Elephant as Person" *Aeon* 24th October 2018.
---. 2019. "Consciousness, Language, and the Possibility of Non-human Personhood: Reflections on Elephants", *Journal of Consciousness Studies*, Volume 26, Numbers 3-4, 2019, pp. 227-251(25)

Rovane, Carol, A. 1997. *The Bounds of Agency. An Essay in Revisionary Metaphysics,* Princetown: Princetown University Press.
---. 2014. "Group Agency and Individualism". *Erkenntnis* 79, pp. 1663-1684.

Rovelli, Carlo. 2017. *Reality Is Not What It Seems.* Penguin Books.

Rowlands, Mark. 2010. *The New Science of the Mind*. The MIT Press.
---. 2013. "Sartre, Consciousness, and Intentionality", *Phenomenology and the Cognitive Sciences*. 12/3:521-36.

Ruiz-Mirazo, K, and Moreno, A. 2004. "Basic Autonomy as a Fundamental Step in the Synthesis of Life" *Artificial Life* 10: 235-259 (2004).

Russell, Bertrand, 1912. *The Problems of Philosophy*, Oxford: Oxford University Press.
---. 1914. "On the Nature of Acquaintance," *Monist*, 24: 161–187.
---. 1921. *The Analysis of Mind*. London: George Allen & Unwin Ltd.
---. 1927. *An Outline of Philosophy.* London: Routledge.

Ryle, Gilbert. 1949. *The Concept of Mind.* Hutchinson.

Salice, A. 2016. "Collective intentionality and the collective person in Max Scheler". In Harald A. Wiltsche & Sonja Rinofner-Kreidl (eds.), *Analytic and Continental Philosophy: Methods and Perspectives. Proceedings of the 37th International Wittgenstein Symposium.* De Gruyter. pp. 277-288.

Salice, A. and K. Miyazono. 2019. "Being one of us. Group Identification, Joint Actions, and Collective Intentionality". *Philosophical Psychology*.

Salice, A. and Zahavi, D. 2016. "Phenomenology of the we; Stein, Walther, Gurwitsch". In J. Kiverstein (ed.) *Routledge Handbook of Social Mind*, London: Routledge.

Sartre, J-P. 1942/1969. *Being and Nothingness.* Translated by Hazel E. Barnes. London: Methuen & Co Ltd.

Sass, L. 1998. 'Schizophrenia, Self-consciousness and the Modern Mind', *Journal of Consciousness Studies* 5: 543-65.

Schaffer, Jonathan, 2010. "Monism: The Priority of the Whole", *Philosophical Review*, 119(1): 31–76

Schopenhauer, Arthur. 1818/9, 1966. *The World as Will and Representation*, translated by E.F.J. Payne, Dover Publications.

Schrödinger, Erwin. 1926. "An Undulatory Theory of the Mechanics of Atoms and Molecules", *Physical Review* 28(6): 1049-1070.
---. 1964. *My view of the World.* Transl. by Cecily Hastings. Cambridge: Cambridge University Press.
---. 1996. *Nature and the Greeks.* New York: Cambridge University Press.

Seager, William. 2012. *Natural Fabrications. Science, Emergence and Consciousness.* Springer.
---. 2014. "Why Physicalism?" in *Mind and Matter* Vol.12, Issue 2, 2014.

Searle, J. R. 1984. *Minds, Brains and Science.* Harvard University Press.
---. 1990. "Collective Intentions and Actions" In Philip R. Cohen Jerry Morgan & Martha Pollack (eds.), *Intentions in Communication.* MIT Press 401-415 (1990)
---. 1992. *The Rediscovery of the Mind.* Cambridge, MA: MIT Press.
---. 1997. *The Mystery of Consciousness.* The New York Review of Books.

Sellars, Wilfred 1963. "Philosophy and the Scientific Image of Man" in *Science, Perception and Reality*, pp.1-40. London: Routledge and Kegan Paul.
---. 1997. *Empiricism and the Philosphy of Mind.* Harvard University Press.

Shani, Itay. 2015. “Cosmopsychism. A Holistic Approach to the Metaphysics of Experience”. *Philosophical Papers* 44/3, 389-43.

Shear, Jonathan. 1999. “Experiential Clarification of the Problem of Self”, in *Models of the Self,* edited by Shaun Gallagher and Jonathan Shear. Imprint Axademic.

Shoemaker, S. 1984. *Identity, Cause, and Mind.* Cambridge. Cambridge University Press.

Slaby, Jan and Heilinger Jan-Christoph. 2013. “Lost in Phenospace: Questioning the Claims of Popular Neurophilosophy”, *Metodo. International Studies in Phenomenology and Philosophy,* Vol. 1, n. 2 (2013).

Smith, Brian Cantwell. 2019. *The Promise of Artificial Intelligence – Reckoning and Judgment.* Cambridge, Massachusetts; The MIT Press.

Sokolowski, Robert. 2000. *Introduction to Phenomenology.* Cambridge: Cambridge University Press.

Spieglberg, Herbert. 1961. On the “I-am-me” experience in childhood and adolescence. *Psychologia* 4: 135-146.

Spinoza, Benedict De. 1677/1996. *Ethics.* London: Penguin Books.

Strawson, G. 2006. “Realistic monism: why physicalism entails panpsychism” *Journal 0f Consciousness Studies.* 13 (10-11), 3-31.
---. 2009. *Selves: An Essay in Revisionary Metaphysics.* Oxford: Claradon Press.
---. 2011. “Radical Self-Awareness” in *Self, No Self?* Mark Siderits, Evan Thompson, and Dan Zahavi (eds.) Oxford: Oxford University Press. Pp.287-8.

Sterckx, Roel. 2019. *Chinese Thought: From Confucius to Cook Ding.* Pelican Books.

Stoljar, S. J. 1973. *Groups and Entities.* Canberra: Australian National University Press.

Tahko, Tuomas. 2012. "In defence of Aristotelian metaphysics" in *Contemporary Aristotelian Metaphysics.* Cambridge University Press.

Tartaglia, J.P.F. 2016. "What is at Stake in Illusionism?" *Journal of Consciousness Studies*, vol. 23(11-12), 236-255.

Tegmark, Max. 2017. *Life 3.0.* Penguin Books.

Thagard, P. 1992. *Conceptual Revolutions.* Princeton: Princeton University Press.

Thompson, Evan. 2007. *Mind in Life.* Harvard University Press.
---. 2015. *Waking, Dreaming, Being.* New York: Columbia University Press.

Tollefsen, Deborah. 2015. *Groups as Agents.* Cambridge: Polity Press.

Tononi, Giulio, 2004. "An Information Integration Theory of Consciousness", *BMC Neuroscience*, 5(November): 42. doi:10.1186/1471-2202-5-42 .
–––. 2008, "Consciousness as Integrated Information: A Provisional Manifesto", *The Biological Bulletin*, 215(3): 216–242. doi:10.2307/25470707
---. 2015. "Integrated Information Theory" *Scholarpedia,* 10 (1), pp. 4164.

Turing, Alan. 1936. "On Computable Numbers" *Proc. of the London Math. Soc*. Pp 230-265, 544-546.
---. 2004/1948. "Intelligent Machinery" In B. Jack Copeland (ed.), *The Essential Turing*, 410-432, Oxford: Oxford University Press.

Van Cleve, James. 1990. “Mind-Dust or Magic? Panpsychism Versus Emergence”. *Philosophical Perspectives,* Vol. 4, Action Theory and Philosophy of Mind (1990). pp. 215-226.

Van Fraassen, Bas. 1980. *The Scientific Image.* Oxford: Oxford University Press.

Van Norden, Bryan. 2017. *Taking Back Philosophy.* Columbia University Press.

Van Inwagen, Peter. 1990. *Material Beings.* Cornell University Press.

Varela, F.J. 1998. “A science of consciousness as if experience mattered” In S. Hameroff, W.Kasznicak, & A.C.Scott, *Towards a science of consciousness II*, Cambridge, MA: MIT Press, 1998.
---. 2000. *El fenomena de la vida.* Santiago, Chile: Dolmen Essayo.

Vendler, Zeno. 1984. *The Matter of Mind.* Oxford: Claradon Press.

Virchow, R. 1978 [1858] *Cellular pathology as based upon physiological and pathological histology*, Special ed. Classics of Medicine Library, Birmingham, Ala.

Walmsley, Joel. 2012. *Mind and Machine.* Palgrave Macmillan.
---. 2015. “Emergence, group judgment and the discursive Dilemma”. *Mind Soc* (2015) 14:185–201.

Walsh, Denis. 2016. “Hylomorphism, foundational materialism, and emergence” at *Biological Identity Conference.* Institute of Philosophy, London. 3-6-2016.

Watanabe, T. 2009. 'From Spiegelberg's "I-am-me" experience to the solipsistic experience'. Paper presented at the International Human Science Research Conference, Molde, Norway.16-22, June 2009.

Weinberg, Steven. 1992. *Dreams of a Final Theory.* New York: Vintage Books.

Whitehead, Alfred North and Russell, Bertrand. 1910, 1912, 1913. *Principia Mathematica*. 3 volumes. Cambridge: Cambridge University Press.

Wittgenstein, L. 1922/1961. *Tractatus Logico-Philosophicus.* Routledge & Kegan Paul Ltd.
---. 1953. *Philosophical Investigations.* Basil Blackwell Ltd.
---. 1969. *On Certainty.* Oxford: Blackwell.

Wright, E. 2008. *The Case for Qualia.* Cambridge: MIT Press.

Wykretowicz, Hubert. 2018. "Husserl's Unnoticed Contribution to Selfhood" in *Constructivist Foundations* Vol. 14, No 1.

Xiaolu Guo, 2017. *Once Upon a Time in the East: A Story of Growing Up,* Chatto & Windus.

Zahavi, Dan. 1999. *Self-Awareness and Alterity*. Northwestern University Press.
---. 2003. "Phenomenology and metaphysics", Published in D. Zahavi, S. Heinamaa & H. Ruin (eds.): *Metaphysics, Facticity, Interpretation.* Contributions to Phenomenology. Kluwer Academic Publishers. Dordrecht-Boston. 2003, 3-22.
---. 2008. *Subjectivity and Selfhood: Investigating the First-Person Perspective.* Cambridge, MA: MIT Press.
---. 2014. *Self & Other: Exploring Subjectivity, Empathy, and Shame.* Oxford: Oxford University Press.

---. 2017. *Husserl's Legacy: Phenomenology, Metaphysics, and Transcendental Philosophy.* Oxford: Oxford University Press.

---. 2019. *Phenomenology – the basics.* Routledge.

---. 2020. "Consciousness and Selfhood: Getting Clearer on for-Me-Ness and Mineness" in U. Kriegel (ed.): *The Oxford Handbook of the Philosophy of Consciousness.* Oxford: Oxford University Press.